Computational Method Design for Study of Protein–Ligand Interactions

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Celine D. Amoreira

aus Frankreich

<u>Promotionskomitee:</u>

Prof. Dr. Kim K. Baldridge (Vorsitz)

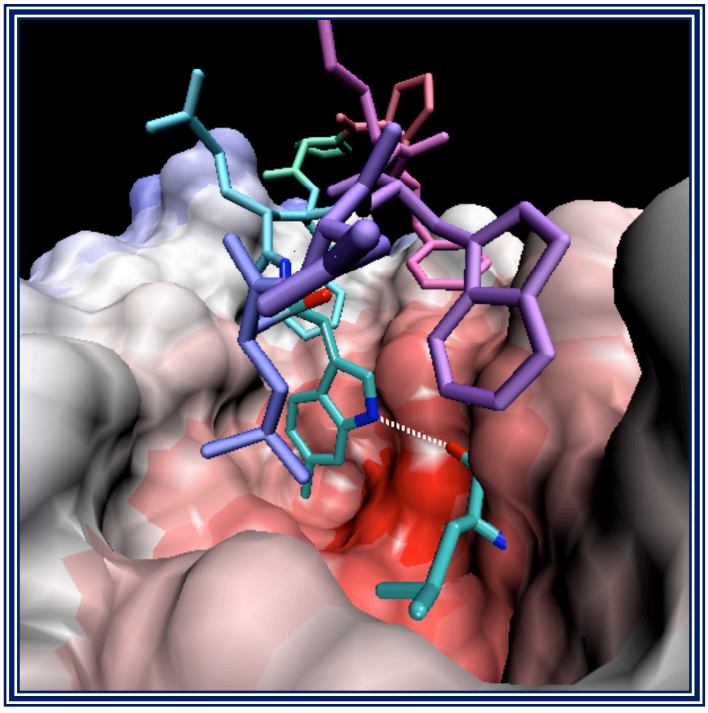
Prof Dr. John A. Robinson

Prof. Dr. Jay S. Siegel

Prof. Dr. Roland. K. O. Sigel

Zürich, 2008

Computational Method Design for Study of Protein-Ligand Interactions



ABSTRACT OF THE DISSERTATION

Computational Method Design for Study of Protein–Ligand Interactions

by

Celine D. Amoreira

Doctor of Philosophy in Chemistry

University of Zürich, 2008

Professor Kim K. Baldridge, Chair

The thesis work presented here, focuses on the conformational methodology for prediction ligand–protein interactions, which are extendable to other complex biological systems. The methodology development involves a hybrid method between Quantum Mechanics and Classical Electrostatic, enabling prediction of binding affinity and analysis of specific interactions between a small molecule and receptor.

The originality of the work consists in a combination of computational methods to estimate binding energy of a molecular complex in terms of steric, electrostatic and molecular orbital interactions. The different contributions to the binding energy are set into a mathematical function, called a scoring function. A geometry analysis is performed on the complex to maximize the interactions and to increase the binding affinity. The design of the scoring function is critical in the ability to realistically rank all new conformations of ligands in the binding site. The quantum chemical computations performed on the small molecule (ligand) involved are performed with the chemistry package GAMESS (General Atomic Molecular Electronic Structure System), enabling better estimation of internal stability and provision of more accurate descriptors, such as atomic partial charges required for the complex electrostatic investigation. The binding energy is calculated using the Adaptative Poisson–Boltzmann Solver (APBS), and involves consideration of both

electrostatic and non-electrostatic contributions to the binding. The electrostatic computations are based on continuum dielectric model to represent the solvent as a continuous medium with mobile ionic concentration.

Due to the complexity of the combined hybrid methodology, we exploited the power of grid computing. Combining the power of the grid with the logical sequential actions of a workflow enables one to consider enlarging the investigations across the combinations of poses of the ligand, investigations of various scoring functions, and variation in details of the computational methodology, for more complete understanding into the nature of the binding of a small molecule into an active site.

ZUSAMMENFASSUNG

Computational Method Design for Study of Protein–Ligand Interactions

von

Celine D. Amoreira

Dr. sc. Nat.

Universität Zürich, 2008

Kim K. Baldridge, Vorsitz

Die vorliegende Dissertation beschreibt eine Methode zur Vorhersage von Ligand–Protein-Wechselwirkungen, die auf komplexe biologische Systeme erweitert werden kann. Es wurde ein Hybridmodell zwischen Quantenmechanik und klassischer Elektrostatik entwickelt, das die Vorhersage der Bindungsaffinität and eine Analyse spezifischer Wechselwirkungen zwischen einem kleinen Molekül und einem Rezeptor ermöglicht.

Die Arbeit befasst sich mit einer Kombination von Berechnungsmethoden zur Bestimmung der Bindungsenergie eines molekularen Komplexes bezüglich Orbital-, elektrostatischer und sterischer Interaktionen. Die einzelnen Beiträge zur Bindungsenergie werden mit einer mathematischen Funktion, der sogenannten scoring-Funktion, ausgedrückt. Eine geometrische Analyse des Komplexes erlaubt eine Maximierung der Wechselwirkungen und eine Erhöhung der Bindungsenergie. Die Form der scoring-Funktion ist entscheidend für eine realistische Bewertung der möglichen Konformationen der Liganden im Bindungszentrum. Die Ligandenstrukturen wurden quantenmechanisch mit Hilfe des Programmes GAMESS (General Atomic Molecular Electronic Structure System) berechnet, wodurch eine bessere Abschätzung ihrer Stabilität und eine genauere Bestimmung weiterer Deskriptoren wie beispielsweise partielle Atomladungen zur Untersuchung elektrostatischer Wechselwirkungen möglich wurden. Die Bindungsenergie, für die

sowohl elektrostatische als auch nicht-elektrostatische Beiträge berücksichtigt wurden, wurde mit dem sogenannten Adaptive Poisson–Boltzmann Solver (APBS) berechnet. Als Grundlage für die elektrostatischen Berechnungen diente das Modell des dielektrischen Kontinuums.

Aufgrund der rechnerischen Anforderungen des Hybridmodells wurde Grid-Computing eingesetzt. Die Kombination von Grid-Computing mit den logischen sequentiellen Schritten des Arbeitsablaufs ermöglicht es, eine grössere Anzahl von Ligandstrukturen und scoring-Funktionen sowie Variationen der Berechnungsmethode zu untersuchen. Dadurch können wir einen besseren Einblick in die Ligand–Protein-Wechselwirkung gewinnen.

Curriculum Vitae

Education

2004-2008	Graduate Student in Computational Chemistry Organic Chemistry Institute – University of Zurich, Switzerland
2002-2003	DESS (Superior Specialized Studies Diploma) "Genie Physiologique et Informatique" Interdisciplinary training in data processing and biotechnology. One year postgraduate degree with honors. University of Poiters - France
2001-2002	MA in biotechnology and computer science with honors. University of Poiters - France
2000-2001	BA in biotechnology and computer science with honors. University of Poiters – France
1998-2000	DEUG majoring in biology. University of Limoges – France

Posters

"Investigation of Protein-Ligand Interactions Using Combined Quantum-Mechanical/Classical and Mechanical/Electrostatics Pipeline Aided by Intelligent Grid-Infrastructure Deployment" *Swiss Chemical Society 2005 – Zürich (Switzerland)*

"Investigation on Protein-Ligand Interactions Using Combined Quantum mechanical/Molecular Mechanics and Electrostatics properties aided by the integrated framework GEMSTONE" *Maria Goeppert-Mayer Interdisciplinary Symposium 2006 – San Diego (United States)*

"Development of Hybrid-Quantum Chemical Methods for Modeling Ligand-Protein Interactions: β-Hairpin protein epitope mimetic inhibitors of the p53-HDM2 complex" *Dorothy Crowfoot Hodgkin Symposium 2007 – Zürich (Switzerland)*

Publications

"An improved version of the DNA methylation database (MethDB)." Amoreira, C.; Hindermann, W.; Grunau, C.. *In Nucleic Acids Research*. 2003. p. 75-77.

"The Computational Chemistry Prototyping Environment. " Baldridge, K.K.*; Greenberg, J.P.; Sudholt, W.; Mock, S.; Altintas, I.; Amoreira, C.; Potier, Y.; Birnbaum, A. Bhatia, K.; Taufer, M.. *Special Issue of the Proceedings of the IEEE on Grid Computing*, 2005, 93, 510-521.

"A Framework for the Design and Reuse of Grid Workflows." Altintas, I.; Birnbaum, A.; Baldridge, K.K.; Sudholt, W.; Miller, M.; Amoreira, C.; Potier, Y.; Ludaescher. *In Scientific Applications of Grid Computing: First International SAG2004* , P. Herrero, M.S.Perez, and V. Robles (Eds): Springer-Verlag, 2005, 3458, 119-132.

"GEMSTONE: Grid-Enabled Molecular Science through Online Networked Environments. " Baldridge, K. K.; Bhatia, K.; Greenberg, J.P.; Stearn, B.; Mock, S.; Sudholt, W.; Krishnan, S.; Bowne, A.; Amoreira, C.; Potier, Y.. Invited paper: *LSGRID Proceedings*, 2005.

"Cluster and Grid Infrastructure for Computational Chemistry and Biochemistry. " Baldridge, K.K.*; Sudholt, W.; Greenberg, J.P.; Amoreira, C.; Potier, Y.; Altintas, I.; Birnbaum, A.; Abramson, D.; Enticott, C.; Slavisa, G.. *In Parallel Computing for Bioinformatics* (Invited Book Chapter), A. Y. Zomaya (Ed.), John Wiiley & Sons, 2006.

"Molecular Mechanics"Amoreira, C.; Baldridge, K. K.. *In McGraw Hill Encyclopedia of Science and Technolgoy*, 10th Ed., Blumel, D., Ed., McGraw Hill Professional, New York, 2006.

"A Flexible Grid Framework for Automatic Protein-Ligand Docking" Abramson, D.; Amoreira, C.; Baldridge, K.K.; Berstis, L.; Kondrick, C.; Peachey, T.. *e-Science and Grid Computing, 2006. Second IEEE International Conference*. Dec. 2006 47-47 DOI 10.1109/E-SCIENCE.2006.261131

"Stereoselectivity and expanded substrate scope of an engineered PLP-dependent aldolase." Seebeck, F.P.; Guainazzi, A.; Amoreira, C.; Baldridge, K. K.; Hilvert, D.. *Angewante. Chemie*. Int. Engl., 2006, 45, 6824,6826

"On the Role of Hydrogen Bonding and Electrostatic Interactions in Binding of p53-Derived Peptides and Peptidomimetic Inhibitors to HDM2 From Theory and Experiment" Anja Grässlin, Celine Amoreira, Kim K. Baldridge and John A. Robinson. *Journal of the American Chemical Society* 2008 (Submited)

Dedicated to my dad, Delphy Amoreira

Acknowledgements

The first person I want to thank is my wonderful advisor, Pr. Kim Baldridge for her invaluable advices and constructive comments, which helped me a lot to focus on my project and contributed to the success of the present research. I am really grateful for all her encouragements, guidance and lovely personality. Thank you for all what she taught me, the trust she put on me, the projects she gave me the opportunity to work on and all the great collaboration she offer me.

The second person I want to thank is Dr. Jerry Greenberg. Thank you Jerry for all the new features you add in QMView to support my project and thank you for always having been here to help me along my thesis even with the time difference with San Diego.

I am grateful to Dr. Wibke Sudholt for her advice and guidance in the early start of the project.

I would like to express my gratitude to the experimentalists for their collaboration on this work: Pr. Don Hilvert, Dr. Florian Seebeck, Pr. John Robinson, Dr. Anja Graeslin. It was interesting and helpful to apply the hybrid method on your systems.

I am very much grateful to the PRAGMA students, Kris, Celia, ranmali, Lynn who contributed to the project along their internship, as well as the heart of this great organization, Cindy Zheng.

Thank you to Phil, Mason and Greg to enable me to use an amazing viz wall thanks to Rocks development. It was really exciting to have a walk inside proteins.

I deeply appreciate the collaboration with Karan, Steve, Brent, Sriram, Jerry and all the members of the Gemstone team. This project was really interesting and provided a useful tool to help me along my research.

I highly appreciate to collaborate with Pr. David Abramson, Slavisa Garic, Tom Peachey, Colin Enticott. I am grateful for all the support you give me when I was stuck with Nimrod. Even if this middleware annoyed me sometimes, it saved me so

much time by automatically launch all my computations. Thank you to Slavisa to have taught me snowboarding during is time in Zürich.

I deeply appreciate the collaboration with Adam Birnbaum, Ilkay Altintas, Mark Miller on Kepler and QMDB projects.

I would express my sincere acknowledgement to my colleagues who created a wonderful social atmosphere, which encouraged everyone to proceed further. They were a real source of help. I thank also all the members of Jay and Nath's groups to have made my time in Zürich so nice.

I am grateful to Simon and Heidi for their help on the german summary.

A special thank to my friends to have support me during this adventure! I really enjoy all the women in science meeting with Fitore, Anne and Becky. Thank you Audrey to have support me remotely and I apologize for the phone bills.

Many thank to my family to have support me and understand my choice.

It is a long list but this reflects a lot of human relationships through collaboration and friendship. Thanks for all of you to have made my work and personal life wonderful along those years in Zürich. Your support was a precious help for me especially when science or personal life was not going well. You are the reason why i enjoy so much my working time and my private life. Thanks!!!!

Abbrevations

APBS	Adaptative Poisson–Boltzmann Solver
CSD	Cambridge Structural Database
CTA	Conformational Tool Analysis
DOF	Degree Of Freedom
HF	Hartree–Fock
GAMESS	General Atomic and Molecular Electronic Structure System
GB	Generalized-Born
MD	Molecular Dynamics
MM	Molecular Mechanics
PB	Poisson-Boltzmann
PBE	Poisson-Boltzmann Equation
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
QM	Quantum mechanics
SASA	Solvent-Accessible Surface area
SAV	Solvent-Accessible Volume
SF	Scoring Function

Table of contents

Table of Figures

List of Tables

1 Introduction and motivations

1.1 Project summary

The project research consists of development and integration of a hybrid method between Quantum Mechanics and Classical Electrostatics for prediction of binding affinity and analysis of interactions between small molecule ligand(s) and corresponding receptor(s). The originality of the work presented here consists of a combination of computational methods design for investigations of molecular complexes in their environment, including understanding of steric, electrostatic, and molecular orbital components. Contributions to the binding energy are set into a mathematical function, called a scoring function. A geometry analysis function is designed and invoked to understand molecular interactions with maximum binding affinity. The design of the scoring function is central to the present work, as the quality of such a function ultimately determines the effectiveness of the method for prediction of protein-ligand interactions. The prediction of structural and binding phenomenon in key research investigations is driving by the scoring function designed to fulfill accurate analysis by Computational methodoly.

The thesis is broken down into i) background and motivation, ii) description of important methodology, iii) hybrid methodology design, iv) research investigations, v) new integrated methodology, vi) conclusions. More specifically, the methods section will provide an overview of computational methods important for this work, and the hybrid methods section will describe in detail the designed methodology based on these computational techniques. The hybrid method is then used in research investigations of two different biological systems to better understand the structural and mechanistic components of these key ligand-macromolecular interactions. Figure 1-1 shows the key steps in such an analysis and provides an overview of scoring function to carry out the analyses using the designed hybrid methodology. The details of these steps will be laid out in the supporting chapters.

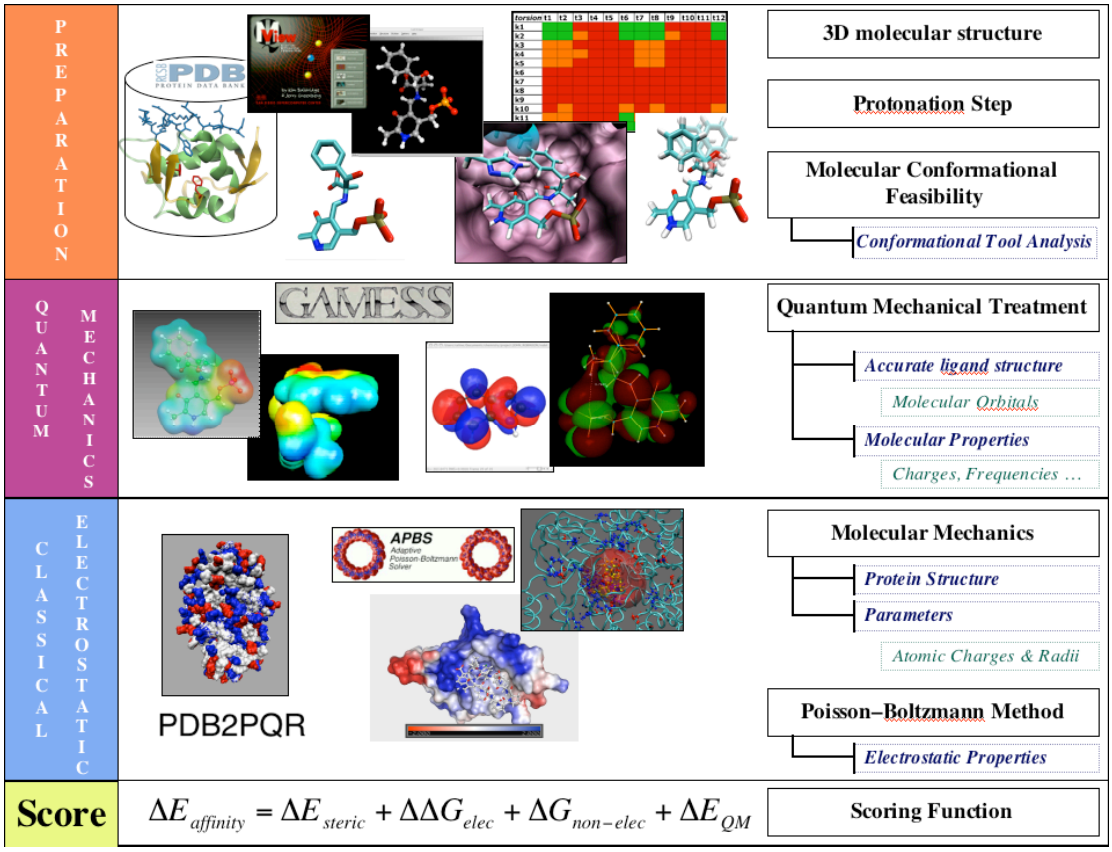


Figure 1-1 Steps involved in the hybrid method to estimate the binding affinity between a small molecule and its targeted receptor.

The scoring function designed to rank the different conformations of the complex is decomposed in several energy components:

- A steric term related to the conformation analysis (ΔE_{steric})
- A term derived from the internal energy of the ligand calculated with Gamess (ΔE_{QM}).
- An electrostatic energy ($\Delta\Delta G_{elec}$) term from APBS calculation with a non-electrostatic term ($\Delta G_{non-elec}$)

1.2 General overview

In the last few years, alerts such as bird flu, human immunodeficiency virus, cancer, and other rare isolated diseases, gather the efforts across fields to elucidate the mechanism of the biological system involved in such diseases, in hopes of avoiding epidemic scares. Such multidisciplinary research involves levels ranging from human body response, to cells, to biomolecules, and to small molecules. The essential asset in therapeutic development strategy is to combine the knowledge of experts at each level of system scale to gain the appropriate knowledge of the responsible target, and how to deal with it. The understanding of chemical/biological processes at a microscopic level contributes to retracing the metabolic pathway. One of the links in this strategy involves understanding the roles played by initiators and/or inhibitors involved in the biological process, and analysis of out how those actors react under environmental stress or genetic modification.

A critical factor is the knowledge of why our immune system may not respond to certain diseases, resulting in increased disorder of biological processes in a way that the body itself is no longer able to deal with the damages. For example, in the case of Alzheimer disease, the accumulation of a small molecule that forms "amyloid plaques," causing brain damage that deprives one of memory. We then ask, why does the body not respond in a positive way to fight against such neurological disorder? The body responds only with known/available processes, the code composed by a combination of only 4 letters. The uniquely individual coding, carry by the DNA, dictates the behavior of our organism. Sometimes the DNA is subjected to irreversible damages and enunciates new rules, which are then followed by cells as a good worker wanting to do the best to keep in harmony with the biological system. When new rules result from a mutation in the code caused by environmental factor, exposure, or genetic predisposition, setting out of control the cell division, cancers can begin to develop. Such activity results in tumor development, and the normal gene carrying the information transforms into an oncogen.

Fortunately, there are guardians[1] of the human genome to prevent such genome mutation, such as the p53 transcription factor that works as a tumor suppressor by

regulated cell cycles. When p53 binds to the DNA, it initiates the processing of repair. However, what happens when a significant mutation manifests itself on chromosome 17, which codes for the p53 protein, or, when an inhibitor molecule blocks the binding process between p53 and DNA? This protein is no longer able to bind the DNA, and cell division is no longer under control. Such a phenomenon opens the door for development of approaches/strategies to ensure safety of the molecular mechanism to fight against tumor invasion. P53 binds in a specific way to the DNA, due to a complementarity between the transactivation domain of p53 and the binding site of the HDM2 receptor. It is relatively easy now to understand how a mutation occurring in p53 that modifies the structure of the recognition domain prevents the interaction between the protein and the DNA, inhibiting in this way the suppressing activity. This illustrates the role played by in molecular recognition.

The human organism is enabled also by a natural p53 inhibitor, like HDM2 [2, 3]. One strategy could consist in synthesizing a peptidomimetic to mimic the domain of p53 involved in the recognition of HDM2 so that the tumor-suppressor is free to activate the DNA repair. From this well-known example, one research target in this thesis will focus on the details of the structure and mechanism of the complex HDM2-peptidomimetic. After isolation of the targeted system in laboratory, the sample is analyzed by X-ray crystallography, NMR, and modeling techniques. The structural and mechanistic analysis provided by computational methodology provides an important bridge to the (bio)chemist. The experimental structures provide the base for theoretical pursuit of understanding of the interactions at an atomic level, and enable more detailed information regarding the reaction process and nature of the interaction between a receptor and its ligand along a disease pathway. In drug design process, for example, research works towards finding appropriate small molecules capable of activating or inhibiting a reaction in the targeted receptor to target a disorder causes by a disease.

Theoretical studies support the knowledge of biochemists and biologists through modeling of a biological system, providing more detailed description of the binding mechanism of a substrate and an enzyme. Enzymatic reactions driven by small molecule(s) represent another class of research, as they are important in many vital biological processes (e.g., protein Kinases). It is crucial to better understand

enzymatic mechanisms in order to act to inhibit, restore, or improve the reaction. From a theoretical point of view, this means combining different levels of analysis with different levels of theory. The protein can be studied with empirical methods and the small molecules can be treated with ab initio based methods. The combination of methods has the potential of providing more accurate assessment of the interaction mechanism between small molecules and macromolecule. However, such methodology development still presents several challenging issues for the researcher, and even to date, there are many improvements that can be made.

Fully understanding the mechanism of action of enzymes is an important step in the use of enzymatic complexes in biotechnological processes. Biological systems can be used in biocatalysis processes, isolated, or in the cell – so called enzyme technology or applied biocatalysis. By addition to the natural enzyme, the major challenge in bioengineering is in the development of new enzymes with predetermined activities. Enzyme biocatalysis processes involve a large field of applications, from the production of processed foods and processed textiles, to fine chemicals and pharmaceuticals. As a recent example of biocatalysis in the pharmaceutical industry, Pfizer researchers developed an enzymatic process to produce a key intermediate in the synthesis of HIV protease inhibitors [4]. Researchers improved the activity and enanthio-selectivity of the enzyme. Evolutionary approach [5] is used to generate new enzymes with different substrate specificity, such as regio-selectivity, enantio-selectivity, thermostability or increasing activities in organic solvent. In the laboratory, site-directed mutagenesis using PCR (Polymerisation Chain Reaction) can initiate substitution of one or more amino acids in a protein. When the mutation targets an active site of an enzyme, the result can turn off the enzymatic activity, thereby improving the rate of the reaction, or by changing the activity of the enzyme. For example, this type of process was used on the binding site of a PLP-racemase dependant enzyme [6]. Racemase activity was modified to an aldolase activity through the reaction process along that the two enzymes share. Aldolase reactions are really attractive in biocatalytic process as they regulate breaking or making of C-C bonds [7]. Such an investigation forms the basis of the second research investigation illustrated in this thesis project.

In general, nature provides significant motivation to focus on specific protein-ligand interactions in life science, drug design development, and enzyme catalysis. Such investigations provide the impetus for development of both experimental and theoretical methodology for our general understanding of challenging aspects of protein-ligand interactions at the macroscopic and microscopic scales. In this thesis, macroscopic and microscopic scales refer to a range of 10^9 times, to jump into the molecular world. The view between the macroscopic and the microscopic, provide different insights for treating the molecule, as collections of atoms or as electrons spinning around atoms, respectively. The investigations are driven by the need to understand the binding mechanism occurring in a binding site of a receptor, and involve the use of a set of tools and different levels of theory to describe the interactions qualitatively and quantitatively in term of sterics, electrostatic/non-electrostatic interactions, and molecular orbital interactions.

1.3 Research project

The general introduction provided the overall direction of this work, based on the collaboration across discipline and fields of science. Along the research work, significant collaborations were enabled with researchers from different horizons, including experimentalists from biochemistry laboratory, computer science experts in grid computing and software development, student interns interested in doing research projects, coinciding with the focused efforts involving the development of methodology to answer key questions at the level of the small molecule/macromolecule interaction interface. Such collaborative interactions contribute to the significance and impact of the research project presented here. As many important and fundamental questions in biology, biochemistry and biophysics can be better understood through investigations on protein-ligand interactions, the goal of the current work is to bring to bear theoretical methods and computational class of interactions. What follows lays the groundwork for this research.

1.3.1 Molecular recognition

Basic aspects of ligand-protein interactions, categorized under the general term *molecular recognition [8, 9]*, are concerned with the specificity as well as the stability of a ligand to bind in a noncovalent way to the binding site. In this point, molecular recognition is central to the development of active substances that might be used as a drug, or a ligand with specialized function. This area of research highly benefits from collaborations between experimentalists and theoreticians, as outlined in the introduction.

In the laboratory chemists, biochemists, and biologists isolate specific molecular species by synthesizing new molecules or characterizing natural products. The laboratory work is an essential step for the development of large databases with structural and energetic information of the new molecular compounds. Experimental research questions involving macromolecular biological complexes are typically directed at understanding the correlation between molecular and functional variations. Biophysical techniques, such as Surface Plasmon Resonance (SPR) or Iso-thermal Titration Calorimetry (ITC), enable one to estimate the binding energy between two molecules, but provide limited insights into the relationship of the ligand and the type of interactions occurring during the binding process. It is really difficult to establish the exact mechanism of action at the molecular level without computational support [10, 11].

Through computational simulation, additional detail is provided on how the system structurally and mechanistically can react in different environments, enabling better characterization of the type of interactions of the system with other molecules. Such information can then be used to improve specific systems, or provide understanding of why a compound is not stable in certain conditions. Each type of analysis method has the potential for providing important feedback to the experimentalists who in turn can utilize the information to improve their synthetic methods, potentially providing increased activity and selectivity of biological constructs. Such a path is often utilized in the process of drug development to find potential candidates for laboratory synthesis. Such synergy between experiment and theory can save considerable money and effort.

Different approaches have been developed that enable proper description of the biological system in the environment in which the biological process occurs in vivo. Computational strategies span from several important classical methods to very detailed quantum mechanical techniques, as well as hybrid QM/Classical methods. The resulting methodology provides the information necessary to formulate the components of the Scoring Function, which is the basis for evaluating a hierarchy of interaction profiles for biological systems. Depending on the choice of computational model, the terms composing the scoring function will have different origins. In the thesis work, the scoring function will take its meaning from the hybrid methodology between classical electrostatics and quantum mechanical methods. The various terms in the scoring function include a component from analysis using Poisson-Boltzmann electrostatic theory [12]. This model enables one to express the binding energy with two primary contributions: polar and non-polar interactions [13]. The apolar contribution is approximated as the cost required immersing the molecule into the biological environment, usually related to the solvent-accessible surface and volume area, plus the attractive solute/solvent interactions. The polar contribution corresponds to the solvation energy and is approximated through description of the electrostatic interaction between the ligand and the protein. Electrostatic interactions represent an important factor in the scoring function because of their strength and long-range nature. In such a strategy, effects of the environment are incorporated through the use of a continuum solvent model [14], which enables characterization of solvent or protein environment through use of a specific dielectric constant. Such a method stands as a common model and is often used in the evaluation of solvation free energies in large biological systems, for reasonable computational cost [15].

An alternative strategy to continuum, or implicit methods, involves an explicit computational consideration of the entire system. In such methods, solvent and protein environment are treated explicitly, and simulations are usually performed using, for example, Monte Carlo methods (for solvent) or Molecular Dynamics, to fully relax the solvent or the entire system, respectively. This type of approach rapidly reaches a high level of complexity and computational cost due to consideration of all atoms in the macromolecular system. However, it can become important to include some level of Molecular Dynamics descriptions to incorporate dynamical information of the system. Such strategies can involve a combination of MD with a continuum

model as well. Research strategies in this thesis work investigate primarily QM/PB hybrid methodology, but also consider influence of molecular dynamics on the researched systems.

1.3.2 Molecular docking: sterics

Computational docking investigations have been recognized as successful aids in the domain of drug design or enzyme recognition. The domain of computer-aided drug design is an active area of research, with the docking procedure handled in different ways. Receptor-based methods build pharmacophores by spatially organizing functional groups to optimize specific favorable interactions at the binding site. Other methods try to optimize the configuration between two constructs, thereby minimizing the energy of binding. Such strategies are broadly defined as '**molecular docking' strategies,** and have been proven to be of great help in drug-discovery [16, 17]. The strategies aim to maximize the interaction between the two molecular constructs through steric compatibility. Algorithms are based on modifying the orientation of the molecules, with the complexity increasing with the degree of flexibility allowed on the model. Usually flexibility is applied to the ligand along different rotatable bonds, perhaps with steric constraints imposed by the position of residues in the binding site. Several docking programs today provide such features, and may additionally include other degrees of complexity by also enabling flexibility at the receptor site.

The present work involves a docking tool based on the flexibility of the ligand. The conformational flexibility was initially incorporated by generation of structural possibilities in a systematic way through rotation of defined dihedral angles. Typically, the most important interactions can be found through dihedral angle flexibility. Generation of possible structures from dihedral rotations, results in a large set of possible conformations. In order to avoid computations on complexes having steric clashes, a checking analysis is necessary. Distances between atoms of the flexible part and the static part are screened to check steric feasibility of the new conformation. Distance checking is also applied between atoms of the new structure and atoms of the receptor, with the aim of eliminating unstable complexes due to

unreasonable intermolecular clashes. The conformational flexibility was subsequently extended to enable additional degrees of freedom flexibility inside the binding site.

1.3.3 Inter- and Intra- molecular interactions: electrostatic and orbital analysis

The importance of electrostatic contributions in the determination of how proteins or substrates interact is well established for some 30 years [18]. Classical electrostatic models based first on Coulomb's law applied to the understanding of such types of interactions, as charge-charge pairs interaction, provides the simplest model. Many subsequent contributions and developments were driven towards improving the accuracy of the theoretical modeling of electrostatic interactions, particularly when such interactions occur in solvent, as it is the case for most of the chemical/biological processes. The computational model becomes more complicated with the treatment of the macromolecules in solvated and salty environment. One way to solve rigorously this complex electrostatic calculation is through the Poisson-Boltzmann (PB) theory. In this theory, the macromolecule is represented explicitly as a collection of fix point charges and atomic radii. The solvent is treated implicitly through a dielectric continuum, and the charge concentration of mobile ions followed Debye-Hückel theory [19]. Significant interest in solving Poisson-Boltzmann equation (PBE) has lead to many improvements in the methodology, enabling one to solve the PBE accurately and quite rapidly [20, 21]. A more detailed description of PBE theory is provided in the methodology section. Successful application of numerically solved PBE has lent itself to successful treatment of substrate binding [22], solvation energy [23], pKa shift [24]and even titration curves [25].

The PBE involves parameters for description of the molecular system of interest. Partial atomic charges and radii are required for the solving of the equation. The common way to assign atomic charges and radii for a protein is through the use of a empirical molecular force field (FF) [26]. Force field parameters are defined empirically or through high-level quantum computations and are not so sensitive to structural modifications. For the protein, the mean structural orientation is imposed through the tertiary structure, so the FF can provide reasonable estimation of charges and radii, but can be quite difficult for representation of the ligand structure. The

ligand is sometimes unknown and also subject to structural constraints to fit into a binding site, both of which can complicate the FF representation, and might required development of specific parameters. The most accurate method for treatment of the ligand remains a description at an electronic level of the microscopic system, using some level of quantum mechanics theory.

1.3.4 Scoring function

The structural complementarity between a small molecule and a receptor is not the only criterion to successful docking search and details of ligand/protein interactions. Steric interactions are also important to account for in the estimation of the affinity between the two molecules. The strength of the binding and/or the mechanism of reaction require determination of energy components from all of the different types of interactions occurring upon complexation. The total binding energy is then the sum of all such energy contributions. The resulting mathematical function representing this summation is called a **Scoring function.** Barney and Kuntz designed the first scoring function based on physics model in 1982 [27]. They developed a geometric approach to macromolecule-ligand interactions for rigid-body docking. Later, flexibility of the ligand was introduced and the scoring function modified to include the additional interactions. Subsequently, many scoring functions have been developed, with some being incorporated in general software for application to biological studies, or, specifically refined.

Scoring functions can be classified into 3 types:

1. Knowledge-based functions: Knowledge-based functions are derived from statistical analysis of intermolecular interactions occurring in protein-ligand structure. The interactions appearing more often than randomly are considered as stabilizing and the interactions appearing less frequently are considered destabilizing. The knowledge-based scoring function is used in the software DrugStore [28]

2. Force Field functions: These functions are based on standard empirical force fields, developed for study of static and dynamic properties of biological systems. Modified version of AMBER and CHARMM force fields have been

used as scoring function to approximate the free energy of binding. For example Autodock and Dock used successfully force field scoring function in their docking program. The score rely on the non-bonded interactions between the protein and the potential binder consisting of electrostatic contribution usually expressed by coulomb's laws, van der Waals terms derived from Lennard-Jones potential and hydrogen bonding.

3. Empirical functions: They are designed to estimate the free energy of binding for a complex protein-ligand when the 3D structure is available. The function use several terms like lipophilic contact energy, ligand-metal binding energy, hydrogen bond and a penalty term applied on DOF of rotation. The coefficients for each term are derived from a regression on numerous complexes with a known binding affinity. ChemScore [29] is using this regression-based scoring functions calibrated with a large set of complexes. The problem with this type of method is the size and diversity of the training set chosen to parameterize the coefficients.

Force field and empirical based functions are energy-based, involving several terms to describe the various types of interactions between two molecular constructs. As most biological reactions occur in solution, contributions from the environment need to be integrated into the evaluation of the binding energy as a solvation energy term. As mentioned earlier, Poisson-Boltzmann theory lead to a good approximation of the solvation contribution to the binding energy.

1.4 Introduction to the resources

In the present work, a hybrid methodology is developed, combining quantum mechanics and classical electrostatic methods to investigate interactions between small molecule(s) and associated target receptors. The intensive nature of the computation and the many levels of processing and analysis drive the technology towards the power of grid computing. The hybrid methodology requires the development and implementation of several tools and programs to perform the entire computation. Moreover, with the ability to use middleware technology, one can build the appropriate bridges between the tools, such as conformational flexibility

enablement, including extraction/alignment of chemical structures into active sites, and analysis of data from generated output and creation of input for subsequent computation with other software. Such repeatable patterns are often the case in investigations of protein-ligand investigations of this type. Combining the power of the grid with the logical sequential actions of a workflow, enable one to enlarge the investigation in many ways, for example, generation of poses of the ligand to investigate, or, process a accuracy hierarchy of computations to understand the nature of the binding through estimation of various scoring functions for the complex.

Today, considerable progress has been made in the development of new technologies and hardware to improve computational efficiency and utility. Such enormous advances have led to correspondingly large impacts in fields of science, such as medicinal chemistry, computational chemistry, and physical bioorganic chemistry. Computational infrastructure development and information technology has offered one of the largest advancements in our society in the last decade. The results are improvements in medical equipment and assisting technology for the doctor in disease diagnosis [30], discovery of new therapeutic treatments through computational supported methodology [31], surgical-enhancement by providing an extra eye for the surgeon enabling operations that before were difficult if not impossible [32]. A great example is the ability to have computation and image-assisted brain surgery, in which robust infrastructure is utilized to combine data sharing, computational analysis, data processing, and acquisition [33-35]. Coupled with computers combined in clusters and in part of multi-level resources of grid infrastructures, enable unprecedented performance in high performance computing (HPC) or high through put computing (HTC). One of the most impressive examples in the grid community is going to be the analysis and storage of data of one of the most exciting experiments running at CERN this year: Large Hadron Collider [36] Computing Grid (LCG) [37]. The data of the biggest physic experiment will be distributed over the biggest worldwide grid infrastructure of computing centers to get enough computational, network and storage resources to exploit the results.

Development of software must follow the hardware evolution and must adapt to the various configurations and operational constraints. Parallel computing already requires software to be split in parts appropriately to the model to be sent to multi-

processors on one or more machines. Adaptation to grid infrastructure may require software to be able to run in heterogeneous environments across different sites. Standards imposed on the adaptation of the software for such integration increase success of such integration across the grid.

In 1969, Len Kleinrock suggested the most common analogy used to describe the grid: "We will probably see the spread of 'computer utilities', which, like present electric and telephone utilities, will service individual homes and offices across the country"

We are not to that point, but we are not so far either. The grid is now used by large communities from different fields and offers a nearly transparent way for users who do not wish to be aware of all the security and heterogeneity issues of the grid. Ian Foster, with coworkers Carl Kesselman and Steve Tuecke, were the first to provide a powerful tool for distributed computing to the scientific community. Foster is, in fact, seen as the "Father of the grid". Globus toolkit [38, 39], a broadly used middleware tool, enables relative ease in managing security, jobs management, and resources. Today, this tool is often used in combination with other grid middleware specialized with other features.

In computational chemistry, advancements in computer hardware and software enable researchers to perform more complex computations [40]. The grid has the potential as an ideal resource to perform large numbers of similar job types that conform to the specifications of the grid type (in terms of length, and resource limitations). In the routine work of the computational chemist, many different software, varieties of format files, and repetitive tasks, are dealt with. Workflow technologies developed in the last 10 years offer superior alternative to integrate the various applications and build links between them [41]. In the proposed work, such grid middleware enablement are used to join software and tools required for the many steps in the procedure for investigating proein-ligand interactions from start to finish with the hybrid techniques being developed.

In the hybrid methodology, visualization also has an important place, as it is the support to check things from the appropriate protonation states, to visualize of properties, such as electrostatic potential mapped onto the molecular surface resulting

from PBE computations. The development of a working environment integrated into an end-user infrastructure [42] has been an invaluable help in the research work. Along the research project proposed here, a framework called GEMSTONE [43] was developed in collaboration with the research group funded and led by Prof. K. Baldridge, involving several computer scientists. This framework integrates several applications necessary to build a work environment with transparent grid accessibility [44].

In addition to multiple resource access and management of distributed computing, the grid community offers tools for distributed parametric modeling. The middleware tool, Nimrod [45, 46], developed in the group of Prof. D. Abramson was chosen. This particular layer in the grid architecture [47, 48] is specialized for parametric computation. The user provides the definition of parameters required for a computational experiment, and the software carries out a computational sweep across all parameter values. There are two different versions of Nimrod, nimrod/G and Nimrod/O. Nimrod/G manages only parameter scans, and Nimrod/O employs one of a variety of algorithms to perform optimization analysis. In the present work, the grid infrastructure supports a computational workflow designed to take advantage of Nimrod/O for analysis of results. More details will be provided in the methodology section.

2 Molecular Computational Methodology

2.1 Introduction

Considerable effort has gone into the study of protein-ligand interactions, resulting in significant tool development and associated software, which ranges over a wide variety of methodology types. This chapter gives a broad overview to the theoretical and computational methods most commonly employed, and upon which developed hybrid methodology was built, as laid out Chapter 1. This chapter will be broken down into the three primary components of the methodology: molecular structure preparation, structure investigation, and choice of scoring function. A fourth section will introduce the grid computing technology and its invaluable support in the present work. And final section will conclude by summarizing the strategy deployed through the use of all techniques and methods.

2.2 Molecular structure preparation

2.2.1 Classical: Molecular Mechanics

Investigation of molecular recognition in protein-ligand remains challenging due to the difference in size of both molecules. It is time-consuming and resource demanding to perform computations at a high level of theory on the whole macromolecule. Region of interest, usually the active site or binding region defining where the interactions occur, can be investigated at an electronic level. The rest of the protein can be described at an atomic level. Empirical based methods are used on the macrosystem and *ab initio* or semi-empirical methods are used to treat the ligand more accurately. In the current work, force field functions are used to prepare the biomolecule in order to add missing heavy atoms on the extracted 3D structure, as well as hydrogen atoms and to optimize the hydrogen network. Parameters stored in the force fields are used to assign partial atomic charges and atomic radii for each atom of the protein. The structure preparation step is important in structure-based methods applied in continuum electrostatic study as the system is defined as a collection of charges and radii at specific coordinates of the atoms.

Molecular Mechanics [26, 49] is an empirical-based method for modeling of large molecular structure with the overall aim to reach energy minima of the molecular system, obeying Newtonian mechanics. This classical method is used in the preparation of the protein. The function used to determine energy and properties of the system contains several terms divided into bonded and non-bonded interactions, as shown in Equation 2-1.

$$E = \sum_{bond} k_b (r - r_0)^2 + \sum_{angle} k_\theta (\theta - \theta_0)^2 + \sum_{improper} k_\theta (\zeta - \theta_0)^2 + \sum_{dihedral} \frac{k_\varphi}{2}\left[1 + \cos(n\varphi + \delta)\right]^2 + \sum_{coul} \frac{q_i q_j}{r_{ij}} + \sum_{vdw} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right]$$

Equation 2-1 Potential energy function

Harmonic functions define the intramolecular interactions of the system represented by bond stretching, angle bending and improper dihedral. Dihedral torsion interactions are represented via a periodic function. Each term contains empirical parameters, the combination of which must be optimized. Non-bonded interactions are governed by charge-charge interactions that obey Coulomb's law, and van der Waals interactions as described by Lennard Jones potential [50]. The set of empirical parameters used in the functions also known as force field are stored in databases that can be embedded into software program like PDB2PQR pipeline used in the present study. Those values are typically derived from a combination of experimental data or/and quantum computations. Several well-known force fields are available, which have been parameterized specifically for a class of molecular systems, and may depend on specific conditions. Force fields depend strongly on the data set used to calibrate them. The pioneer in force field development is Allinger with MM force field optimized for small molecules. Two of the most popular force fields remain AMBER and CHARMM. Based on the same empirical potential, they differ in the way the parameters are derived using different charge model, different van der Waals

for example. A list of some of available force fields is given in Table 2-1 with their characteristics.

FORCE FIELD	SPECIFICITY DATA SET	OPTIMIZED FOR	RADII	CHARGE
MM2/3	Heat of formation Vibrational frequencies	small molecules	Hill potential	Dipole-dipole to the bond
AMBER99	From OPLS	Protein, nucleic, peptides	Lennard Jones Potential	Fixed partial charge center on atom (RESP)
CHARMM27	Experimental data	Small molecules to solvated biomolecules		Ab initio atom-centered charges (RESP)
OPLS [51]	Liquid simulation	Small molecules	Van der Walls	Ab initio atom-centered charges (RESP)
PARSE	Derived from other FF	Small molecule Macromolecule Hydration energy	Lennard Jones	Ab initio charges fitting potential

Table 2-1 List of Force fields with their characteristics

Pertinent to the present study, PARSE [52] force field is recognized as appropriate for biological systems studied with dielectric continuum model, and is particularly appropriate for reproducing solvation energy in electrostatic investigations. The atomic charges and radii are optimized for Finite Difference Poisson-Boltzmann (FDPB). They are derived from already existing sets of parameters for hydration free energies, and from optimizing parameters from other force fields or *ab initio*. The

atomic charges are defined by *ab initio* methods to fit the potential electrostatic, and atomic radii are derived from Pauli van der Waals radii and scaled in certain cases (e.g., hydrogen atoms are set to 1.0 instead to 1.2).

2.2.2 Quantum Mechanical methods

At a more fundamental level, quantum mechanical (QM) methods enable detailed understanding of the behavior of systems at atomic scale. This level of theory is required here to estimate accurate properties of the binding region, such the partial atomic charges. Quantum Mechanics arose in early twenty's century with the wave-particle duality theory and the energy quantization of waves. Based on Hartree-Fock (HF) theory **[53, 54]**, the main goal of QM is to approximate the Schrödinger equation. This fundamental equation, composed of an operator acting upon a particular state of the system described by the wave function, returns an observable property of the system. Solving the equation exactly for any system is only possible a system with a single-electron. Several numerical assumptions are required to solve the more complex systems and provide an accurate description of the electronic structure of a molecule. Unfortunately, the QM approach is not efficiently adaptable for treatment of large systems like proteins. Solving the Schrodinger equation for such amount of atoms will be time consuming and CPU demanding. Therefore, empirical methods are used or combined with QM. In the proposed work, the hybrid method is treating the large biomolecule with classical methods as described previously, whereas some fragments of the complex and especially the ligand require accurate description. QM methods are applied on fragments involved in the molecular recognition between the ligand and its targeting receptor. The interactions can be of different kind as hydrogen bonding, van der Walls interaction or charge-charge interaction. The accuracy of the charge determination in the hybrid method is fundamental in the present study as they are directly involved in the electrostatic calculation. The empirical force fields do not provide really accurate atomic charges, so using the *ab initio* quantum chemistry program GAMESS [55] will provide a high level of accuracy to determine the atomic charges of fragments from macromolecular complex.

2.2.2.1 QM Levels of Theory

The Schrödinger equation can be approximated by mainly two different approaches to determine the position of the electrons and nuclei of the system and their properties. First, the Hartree–Fock (HF) theory is based on the many-body electronic wave function and start with a set of approximate molecular orbital. The second approach is the Density Functional Theory (DFT) that tackles the problem by simplifying the many-electron wave functions by using the electronic density as basic variable.

Hartree–Fock theory is based on the description of the molecular orbital of the systems and on the fact that the energy approximate of the system is overestimated. This assumption corresponding to the variational principle leads to a self–consistent field method to reach the minimum energy of the system over iterations. The method neglect electron correlation so post-Hartree–Fock methods are employed to correct this lack. Møller–Plesset [56] theory is an alternative to add correlation energy by perturbing the system. This Post–Hartree–Fock method costs in CPU time and the fee increase with the size of the system studied.

DFT is not a truly *ab initio* method but more a semi-empirical approach using parameters derived from experiments. It is a cheap solution in term of resource and time demand to get accurate description of molecular system. It is often use in QM/MM methods. Using hybrid functional has developed methods combining both approaches, Hartree–Fock and DFT. That hybrid methods use the exchange functional from HF in addition of the exchange-correlation defined in DFT. B3LYP is a hybrid functional use to optimize the geometry of the ligand in the current work. It stands for Becke for the exchange part, Lee, Yang and Parr for the correlation functionals.

Recently new contributions in the field of density functionals enable more accurate description of dispersion. One such contribution are the MO6[57] series functionals, including a certain amount of exchange correlation from Hartree–Fock to the exchange correlation from DFT. The different MO6 functionals differ by the percent of exchange HF that is specific for certain models as for example computation on transition metal will require functional without HF exchange. The last project will bring some work comparison between DFT using MO6 functionals and Møller–Plesset (MP2) level of theory.

In the proposed work, the calculation on the ligand was carried out using DFT methods as there are cheaper in term of CPU cost and give accurate results. So the computation could gain in accuracy by choosing higher basis set in order to calculate accurate properties like the charges required for the electrostatic study later. The Basis set employs is a double zeta valence increased by 2 d- and 1 p- additional polarization functions and diffusion.

2.2.2.2 Basis sets

Basis sets are mathematical functions used to describe molecular orbitals in which the electrons are confined. The geometry optimization of a molecular structure depends directly on the description of the orbitals, which drive the spatial orientation of covalent bond into the molecule. The molecular orbitals are defined as a linear combination of atomic orbitals. The number of basis functions depends of the choice of the basis set and rises also with the size of the system studied. Minimal basis set such Slater–Type Orbitals give a poor description of the molecular orbital by using a small number of basis functions and describing in the same functions the core and valence shells. Since the reaction occur in the valence shell it is important to have flexibility in the space containing the valence electrons. Extension of the shell space gives the electron the opportunity to adjust on the molecular environment. The valence shell can be partitioned into an inner, more compact layer, and a more diffuse outer shell. This split valence basis set, introduced by Pople contributes to a more realistic electron distribution. The computational requirements increase with the size of the basis set, and therefore, the size of the molecular system under consideration. It is therefore important to make judicious choices of basis set for the molecular description. The accuracy of the basis set can be reflected on the prediction of structure and properties of the molecular system under consideration.

The basis set can be extented by adding polarization functions. This functionality adds basis functions to specific shells. For example p functions will be added to the s valence shell, d functions to p valence shell. Polarization functions increase the valence shell space by modifying its shape and so are of great interest in the study of interaction between molecules. Lone pairs or electrons in virtual orbital are the first to react with the molecular environment. The electronic cloud of such species situated

far from the nucleus is hard to describe accurately without increasing the basis set. Another improvement in the basis set is the functionality to add diffusion functions to the valence shell. Diffuse functions are shallow Gaussian functions that increase the size of the orbital at far distances from the nucleus. This extension is a key feature when studying interaction like hydrogen bonds that occur in the active site of an enzyme.

The accuracy of the molecular orbital description depends on the number, type, and extent of basis functionalities. The choice of the basis set is an important factor in the method proposed here, as the goal is to investigate on molecular interaction. Comparisons of an experimentally known prototype system across a series of basis sets facilitate accurate choice of basis set for the unknown system.

2.2.2.3 Partial atomic charges

An important contribution to the binding energy of protein/ligand interaction is the electrostatic energy resulting from charge-charge interactions. Classical study of electrostatic interaction requires set of parameters defined in the molecular force field. Typically, the electrostatic energy is computed using Coulomb's law, employing partial atomic charges that are derived from high-level quantum mechanical calculations. Partial atomic charges, if accurately represented, can provide information about the reactivity of molecules, as well as the molecular structure and stability. Atomic charges are not physically measurable observables, and computational prediction of charges is complicated by the fact that partitioning of electron density over atoms is not well defined or unique. This property can be mathematically computed to reproduce measurable quantities, such as dipole or electronegativity. Several approaches have been developed to derive the atomic centered point charge from quantum chemical computations, ranging from rather crude methods of physical partitioning [58], to more accurate models involving properties of the atoms, such as electronegativity and electron density zero flux surfaces [59-62]. A summary of the most common methods is presented in Table 2-2.

	METHODS	**DEPENDANCY**
Mulliken [58]	Population analysis of wavefunction	Strongly dependant on the choice of the basis set and the structure geometry
NPA [60]	Population analysis of wavefunction	Small dependance on the choice of the basis set but strongly dependant on the geometry of the moelcule
AIM [59]	Partionning of electron density distribution	Require higher basis set but not too sensible of the structural arrangement
RESP [61]	Charge derived from electrostatic potential	No real dependance on the basis set or change in the molecular structure
ChElPG [62]	Charge derived from electrostatic potential	No real dependance on the basis set or change in the molecular structure

Table 2-2 Partial charge determination schemes, and associated characteristics of each model

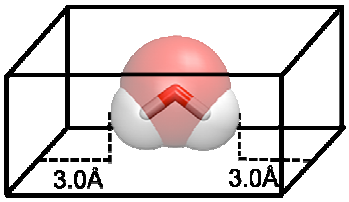


Figure 2-1 CHarge from ELectrostatic Potential using a Grid based method (CHELPG)

A more accurate, and relatively common method used in biochemical computations is the CHELPG [62] schema, derived from the initial CHELP [63] algorithm. In this method, the charges between the van der Waals surface plus an extended 3Å region is fit to reproduce the electrostatic potential for the molecule (Figure 2-1). The electrostatic potential is determined from the wavefunction as described in Equation 2-2,

$$V(r) = \sum_A \frac{Z_A}{|r - R_A|} - \sum_{\mu,\nu} P_{\mu,\nu} \int \frac{\varphi_\mu(r')\varphi_\nu(r')}{|r - r'|} dr'$$

Equation 2-2

Z_A is the nuclear charge on atom A centered at R_A, $P_{\mu\nu}$ is the one-electron reduced density matrix element between the atomic basis functions φ_μ and φ_ν. The fitting method involves carrying out a least-squares fit of the classical electrostatic potential

$$E\ (r, Q_A) = \sum \frac{Q_A}{|r - R_A|}$$

Equation 2-3

with Q_A the atomic centered partial charge, and the quantum mechanical electrostatic potential,

$$F(Q_A) = \sum \left[V(r) - E(r,Q_A)\right]^2$$

Equation 2-4

Minimized under the constraint:

$$\sum Q_A = Q_{tot}$$

Equation 2-5

Improvement of the CHELPG method by Breneman and Wiberg involved a decrease in the orientation dependency of the molecule and an increase in the sensibility of the electronic redistribution during internal rotations. This method has been proved to give comparable results with cheaper basis set than the RESP method that is used to parameterize AMBER force field.

In general, the introduction of the Molecular Electrostatic Potential (MEP) in 1969 and the Molecular Polarization function in 1976 contributed a lot in the development of the Continuum solvent model and the Polarizable Continuum Model (PCM). The model consists of describing the effect of non–convalent interactions between the solute and an average distribution of mobile molecules representing the solvent.

2.2.2.4 Environmental effects: Polarizable Continuum Model

As most all (bio)chemical reactions occur in some kind of environment, any electronic structural computational analysis, including geometric and property effects, needs to be carried out in the environment. The most efficient way of including the solvent in the computation is through a dielectric continuum model. Dielectric continuum models, first developed by Born in 1920 [64] began with the estimation of the solvation energy for an ion in a constant dielectric. This was followed by Debye–Hückel in 1923 [19] who applied the method to an ionic solution. The continuum model was extended by Kirkwood in 1934 [65] to include several charges surrounded by a spherical cavity. Onsager contributed to the model by calculation of the solvation free energy for the dipole in spherical and ellipsoidal cavities [66]. These initial developments enabled introduction into the Self–Consistence Reaction field (SCRF) model of QM based electronic structure methods by including the effect of the solvent into the Hamiltonian of the solute [67]. Through the self-consistent iteration, the polarization induced by the solvent is taken into account as well as the polarization back onto the solvent. In the last 10 years, continuum solvation methods in QM have become more and more sophisticated [14]. Different approaches have been proposed to improve the solvent cavity. For example, Rivail suggested the distributed multipole expansion for spherical or ellipsoidal cavity shape [68]. Tomasi and coworkers [69, 70] introduced their Polarization Continuum Model (PCM), which is implemented in several QM software, including GAMESS. Baldridge and Klamt introduced the Conductor-Like Screening QM continuum model in GAMESS in 1997 [71], adding key corrections to the standard continuum model, including corrections for embedded cavity surface elements, outlying charge, and surface charge representation [72]. Most popularly used continuum models now include these corrections.

For the present research, the PCM solvent model enables modeling of the environment in which the ligand is subjected when it is complex within the receptor. Using a dielectric of 4, it is a reasonable value to reproduce the interior of the protein [73, 74].

2.2.2.5 Geometry optimization under constraint

The computation to perform geometry optimization can be done on the entire system. A full geometry optimization gives the ground state energy representing the minimum energy of the system in its more stable state. Geometry optimization can be performed with constraints applied to the system. The present study involves the crystal structure of the macromolecular complex in its bounded form. Only the hydrogen atoms of the molecule will be optimized and all the heavy atom positions are frozen during the optimization process. A lot of conformations of the molecule are generated in the preparation step and the goal is to keep this particular atomic arrangement to investigate the fitting inside the binding site. Molecular recognition involves specific hydrogen bonds and charge-charge interaction, so it is important to optimize the position of the hydrogen atoms.

2.3 Molecular Conformational Feasibility in Protein Systems

2.3.1 Molecular docking introduction

Producing viable pharmaceuticals in laboratory can be extremely time-consuming and expensive. Significant efforts have been extended to develop automated methods to cut both the time as well as the expense. Computational methods are often invoked to accelerate the laboratory investigation. There are actually two main categories of such methods used in drug discovery:

- *De novo design [75]:* In this process, a molecule is built up from scratch, typically by way of a fragment approach. This method employs a library of known fragments and assembles different sequences of fragments by looking the complementarity between the fragment selected and the part of the binding site that the fragment is supposed to interact with and bind optimally. The fragments are finally linked together to provide an optimal binding ligand for the targeted receptor.

- *Docking:* Typically, a library of known molecules from a database are systematically replaced into the binding site of the receptor of interest and each binding affinity estimated by application of a Scoring Function (SF).

In the present work, a docking method is employed, consisting of two main components. The first component involves the search over possible ligand geometries to analyze best fit in the pocket of the protein. The second component is the application of a Scoring Function (SF) to rank each new conformation of ligand. This chapter will focus on the first component, discussing the importance of generating new conformations of the ligand in the receptor and an analysis of the feasibility of the new conformation in terms of steric compatibility.

A wide range of docking programs have been developed over the last past decade [76, 77]. These methods primarily involve two different molecular recognition approaches for description of the system of interest, Rigid Receptor Docking and Induced Fit.

2.3.1.1 Rigid receptor docking

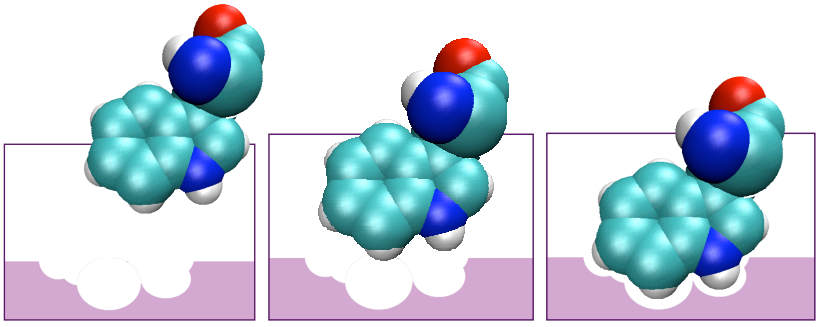


Figure 2-2 Lock-and-key model described by Emyl Fischer in 1894

In this approach, the receptor is treated statically with all internal degrees of freedom defined in the ligand. Rigid receptor docking is based on the Lock-and-Key models (Figure 2-2) described by Emyl Fischer 100 years ago[78]. In this model, the search relies on perfect geometry complementarity between a receptor and a potential ligand candidate. The receptor and the ligand are treated as rigid components in the docking

procedure. The ligand is allowed to move only along 3 translational (x, y, z) degrees of freedom, 3 rotation degrees of freedom, and internal rotations along torsion angles of rotatable bonds of the small molecule.

Considerable software development efforts used the rigid model at the beginning of the docking development applications. Simplification in the computation was important at that time because of the hardware limitations and also because it was desirable to be able to screen a large library of compounds in an reasonable amount of time. This model is still used quite a bit in docking programs due the simplification in generating new conformations in a reasonable amount of time. The rigid-receptor docking model is not highly accurate, however if the structure of the receptor used in the investigation is the one induced by the ligand, then any inaccuracy of the assumptions is decreased. In addition, X-ray diffraction or NMR techniques are able to provide accurate structures of receptors bound to their ligand, thereby facilitating the process.

2.3.1.2 Induced fit

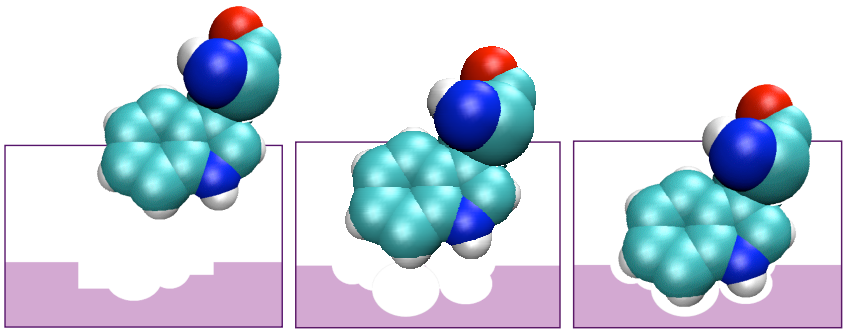


Figure 2-3 Induced fit model describes by Daniel Koshland in 1958

This model was introduced by Daniel Koshland in 1958, and is based on the idea of an induced fit of the ligand into the protein[79]. It is well established now that certain receptors in their apo-form, when they are not interacting with their ligand, adopt a completely different shape than when they are bound (Figure 2-3). As an example, the

HDM2 receptor adopts a totally different structure when it is bound to its inhibitor, the trans-activation domain of p53 [80]. This dynamic behavior illustrates the importance of enabling some level of flexibility in the receptor in the binding region. The induced-fit model is more realistic than the rigid docking model, as in vivo, the receptor is responding to the presence of the ligand, which in turn modifies its chemical and structural environment. Additionally, a mutual adaptation between the ligand and the protein is required to initiate the interaction. From the computational perspective, the induce-fit model implemented in a docking tool increases significantly the complexity of the computation and the time to dock a receptor with a new inhibitor molecule. Further study of the dynamic process of binding between a receptor and a ligand involves simulation techniques involving Molecular Dynamic, further complicating the computational process.

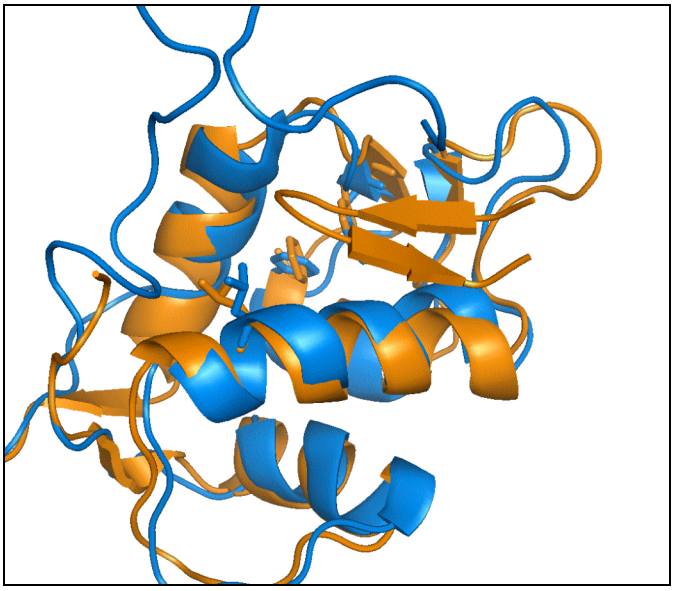


Figure 2-4 Apo-form of HDM2 protein (1Z1M) in blue and HDM2 structure bind to beta-hairpin peptidomimetic (2AXI) in orange.

2.3.1.3 Geometry search

The geometry investigation in docking study is driven by several strategies. The overall aim is the determination of a certain degree of complementarity between a receptor, usually a protein, and a small molecule. The first step is the decision of what flexibility to enable in the biological system. As mentioned previously, the flexibility can be provided to the ligand by only investigating different orientations into the receptor. The search algorithm used to carry out the flexibility mapping of the ligand usually follows a systematic search method along which all the new conformations of the ligand generated for each type of DOF is considered. Nonsystematic method will consider a sampling of all of the possibilities. There is a random way to generate conformations and study them. This method is based on a stochastic analysis that will perform one rotation at a time, with the estimation of the impact of the motion towards the location of the local minima. Finally, deterministic search algorithm involves the use of the current generated state to generate the next state in order to orient the search to a local minimum. This method is time consuming and has the possibility of driving the search in a wrong direction.

2.3.2 Conformational Tool Analysis (CTA) develop internally

In the present study the flexibility is applied on the ligand and on some side chains of the receptor. The structural changes are made along specific rotatable bonds. The goal is to define a geometrical complementarity between the ligand and the receptor. In order to check the feasibility of the generated structures a control on the geometry of the molecule itself and a control inside the receptor are performed. These preliminary checks avoid unnecessary computations in determining the binding energy.

2.3.2.1 Conformation generation

To probe the conformational flexibility in the area in and around the binding pocket, modifications are applied to the ligand or to the receptor side chains using some preferred metric. As discussed, options for such metrics include torsional rotation, full body rotation, translation, or full flexibility. In the present work, investigations along specified torsion angles were first pursued, and all structures generated by

rotation through these torsions is generated and subsequently docked into the active site. The granularity of the rotation must be tested and refined with smaller steps around the region providing any additional room for flexibility. This method is based on the incremental construction algorithm proposed by DesJarlais and Kunz in 1986 [81, 82]. One can consider the creation of a combinatorial library for such a flexibility mapping, which gives an overall picture within the active site.

2.3.2.2 Conformational analysis

The large number of structural combinations generated in the creation of a combinatorial library increases with the number of rotatable bonds considered as well as the step size to scan dihedral angles. However, not all of the generated structural rearrangements will be stable, due to, for example, steric clashing between atoms of the flexible part of the ligand and atoms of either another internal part of the ligand, or, with atoms of the receptor pocket. Therefore, some conformational checking must be performed to eliminate these as possibilities for binding.

Various criteria are chosen to estimate the minimal distance between atoms necessary to preserve stability of the complex. Geometric constraints based on minimum distance avoid steric clash between atoms. They can be checked based on their van der Waals radii and an overlap factor. The overlap factor represents the percentage of overlap allowed between two atomic centers and the sum of their van der Waals radii. This factor is typically set between 0.7 and 0.8. The advantage of using an overlap factor is the independence on atom type, and so in the case of hydrogen bond, the two closed atoms involved in a noncovalent interaction will not be detected as a bad steric contact. Other restrictive criteria may also be selected, such as atomic radii, number of clashes, or fixed minimum distance between atoms.

$$\sqrt{(x_S - x_F)^2 + (y_S - y_F)^2 + (z_S - z_F)^2} \geq \alpha \int_S \int_F (r_S^{vdw} + r_F^{vdw})$$

Equation 2-6

Equation 2-6 expresses the geometric constraints apply on atoms part of the static part of the complex and flexible atoms, in order to check the amount of overlap between van der Waals radii of each atom. The distance between atoms from the static part and the new positioning of atoms of the mobile part of the molecule (LHS) has to be superior to the sum of van der Waals radii of the two atoms scaled by a factor managing the amount of overlap (RHS), where S is static and F is Flexible, respectively, α represents the overlap factor equal to 0.7.

Once checked, the remaining ligand structures are rigorously replaced into the receptor to estimate the degree of steric complementarity. All the distances between atoms from the receptor and atoms from the ligand are screened and checked using minimum distance constraints. Additional intermolecular contacts between the ligand and the receptor can also be considered at this point, since important noncovalent interactions will occur that also can contribute significantly to the binding affinity.

The conformational analysis based on geometrical properties need to be completed by more investigations on molecular recognition. Then a scoring function is used to estimate the binding affinity between the macromolecule and each poses of the ligand. The result of the geometry analysis is usually translated as a factor to rank the different structures generated and this factor is integrated into the scoring function. In the work, this factor will not appear directly in the final designed scoring function as the choice was made to not investigate on the unstable structures.

2.4 Scoring functions methods

2.4.1 General overview

The geometrical analysis of the ligand and receptor needs to be supplemented by additional charge and orbital interactions to fully characterize the molecular recognition, including all interactions stabilizing the complex. Additional intermolecular interactions include van der Waals interactions, electrostatic interactions and directed interactions such as hydrogen bonds, halogen bonds, or disulfide bridges[83]. It is also important to include the solvation/desolvation

contributions to the complexation of the protein and associated ligand. The solvation/desolvation energy contribution represents the work required to bring each pieces of the complex from one environment to another, through gain or loss of entropy and disruption of solvent environment. If portions of the molecular surface are exposed to polar and/or non-polar groups, solvent reorganization may occur. The non-polar groups tend to bury inside the core protein causing a release of solvent molecules and therefore a corresponding rise of entropy. Consequently, the system gains in stability under hydrophobic influences [84-86]. This process drives the folding of most of the proteins and increase their stability [87].

All interactions involved in the complexation process are represented in a mathematical function used to score the different conformations of the complex. The design of the scoring function can range from extremely basic to rather sophisticated. Experimentally, binding affinity is quantified through measurement of the equilibrium constant upon complexation. Biophysical techniques such as ITC are used to measure the binding affinity and the corresponding enthalpy change of the system. From those values the Gibbs free energy and change in entropy can be derived:

$$\boxed{\Delta G_{bind} = -RT\ln K_a = \Delta H - T\Delta S}$$

Equation 2-7

Equation 2-7 expresses the change in Gibbs free energy (ΔG) between two different states of the system, where R is the gas constant, T is the temperature, K_a is the equilibrium constant, ΔH is the change in enthalpy and ΔS is the change in entropy.

Despite advances in computational methods, consistency with experimental results is not reached yet. A relative estimation of the binding affinity can be calculated through the scoring function, to compare with the experimental value. The scoring function is designed by following the thermodynamic circle shown in Figure 2-5. The free energy of binding can be expressed in a simplified way by the difference between the bound and unbound form of the system, as:

$$\Delta G_{bind} = G_{complex} - G_{protein} - G_{ligand}$$

Equation 2-8

Equation 2-8 relates the binding energy of a complex protein-ligand expressed as a difference between the free energy of the bound state, the protein and the ligand.

The energy required to move the molecules from vacuum into solvent contributes mainly in the binding energy. The transfer modifies the electrostatics of the systems due to polarization. Hydrophobic residues onto the surface of the protein will be driven in the core of the protein by hydrophobic effect in contact of solvent. An exclusion phenomenon occurs with the solvent trying to minimize the surface in contact with the hydrophobic amino acid. In order to reflect those interactions, dynamics investigations are required. The shape of the receptor upon binding can be taken and kept constant for this step. This assumption truncates the calculated free binding energy of an important entropic contribution. The solvation energy corresponds to step 2 and 4 on the thermodynamic circle (Figure 2-5). It includes the difference between the electrostatic energy of the molecules into solvent and in a uniform environment, called reference state. This polar contribution to the free solvation energy of a molecule can be calculated by solving PBE with a continuum solvent model.

$$\Delta\Delta G^{polar} = \Delta G^{elec} - \Delta G^{ref}$$

Equation 2-9

Equation 2-9 is the expression of the polar contribution to the free solvation energy of a molecule estimated in term of energy transfer from gas phase to solvent.

When the molecules are transferred into the aqueous medium, a cavity is form into the solvent. The formation of the cavity disrupts solvent molecules organization by breaking hydrogen bond. The cavity creation is described with solvent-accessible surface area of the molecule scaled by solvent surface tension term and solvent-accessible volume of the solute weighted by the pressure of the solvent. Dispersive interactions are generated at the cavity interface to the binding energy and represent the apolar or non-electrostatic contribution (G^{apolar}) to the solvation energy.

Another contribution of the binding energy is the coulombic energy (G^{coul}) corresponding to step 3 of the thermodynamic circle. Upon binding the coulomb electrostatic energy represent all pairwise interactions between all the atoms in a uniform dielectric.

Each contribution to the free binding energy is calculated for each new complex generated and the scoring function described above is applied.

$$\boxed{\Delta G^{bind} = \Delta\Delta G^{polar} + \Delta G^{coul} + \Delta G^{apolar}}$$

Equation 2-10

Equation 2-10 described the scoring function decomposed into a polar and apolar contribution to the solvation energy and coulomb pairwise interactions.

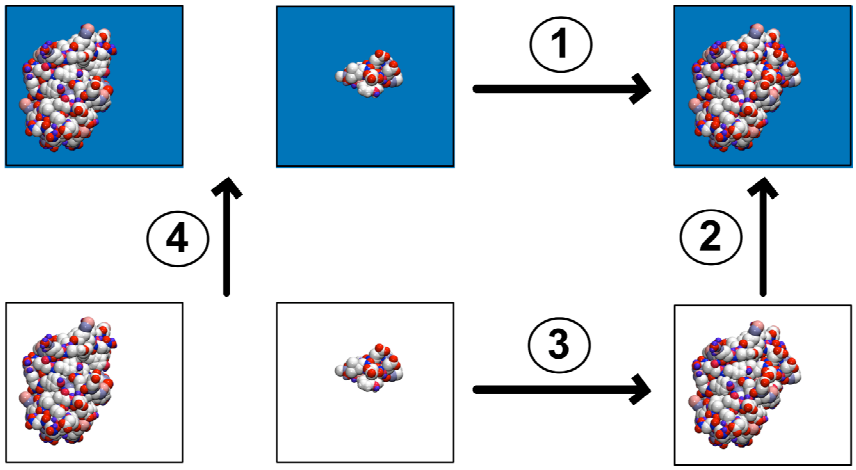


Figure 2-5 Free energy calculation for a biomolecular system.

The blue background in Figure 2-5, indicates a solvated environment (high external dielectric constant) with ionic strength and the white background corresponds to a reference state environment (low external dielectric constant identical to the internal dielectric of the biomolecule). The arrow number 1 indicates the binding of the two molecules in solution. Arrow 3 indicates the binding of the two molecules in the reference state (same internal and external dielectric constant). Arrow 2 corresponds to the solvation energy for the biomolecular complex. Arrow 4 indicates the solvation energy for the isolated biomolecular components.

2.4.2 A continuum approach: Poisson Boltzmann theory

Computational methods to study electrostatic interactions in biomolecular systems can be classified in two main categories based on the representation of the solvent. The explicit solvent models simulate the electrostatics by describing all the molecules of the system including each solvent molecule as well as the ions. Although this model provide a lot of details in the interaction between the protein and the solvent molecules, it remains really demanding in computer power and in CPU time if one considers all DOF for each solvent molecule. The most popular computational

electrostatic approach is based on an implicit representation of solvent and salt through a continuum model. The medium is described by a dielectric constant and the counter ions as an average distribution of charges. This approach was successfully used in the elucidation of electrostatic properties complex [11]. Table 2-3 draws advantages/disadvantages between explicit and implicit solvent model. In the present work the continuum electrostatic method was choose to study protein system in solution as it is cheaper in term of computation time and resources. It is a good approach to estimate protein/ligand binding energies with the contribution of the solvent [88] and the long-range electrostatic interactions are fully taking in account in the whole system. The most current method to describe electrostatic potential and distribution of mobile ions around the molecules in solution is based on Poisson–Boltzmann theory.

	ADVANTAGE	**DRAWBACK**
Implicit environment model	No cutoff (long rang interaction take in account) No solvent sampling configuration errors Not limited by large number of solvent molecule	No information on properties derived by running simulation Separation of $\Delta H/T\Delta S$ difficult
Explicit environment model	Generality of the methods solvent-mediated biomolecular interactions details good properties estimation (h-bond number)	Cutoff Entropy dependant (hard to determine)

Table 2-3 List of advantages and disadvantages between explicit solvent model and continuum approach

Poisson-Boltzmann model, introduced at the beginning of last century by Gouy[89] and Chapman[90], is applied to describe electrostatic properties of molecules in a continuous medium. Debye and Hückel[19] formulated PBE for calculation of the free energy of ionic systems. However, still an analytical solution could be performed only for simple geometry and symmetric ionic repartition [65]. Limited attention was given to this area again until the 80's, due to not only insufficient computer resources, but also limited accuracy in molecular surface shape description. Improvement in the description of the interface between solvent/solute for electrostatic continuum models bring better estimations in molecular recognition in the study protein/protein interfaces in solution [91]. The most important contribution towards the development

of the PBE as a standard method for study of electrostatic effects was the grid-based method define by Warwicker and Watson [20] in 1982. They developed a finite difference approach to calculate the electrostatic potential of a non-regular shape protein. With these developments, classical electrostatic theory using the PBE was reintroduced for study of protein structural stability, molecular recognition, and enzyme catalysis. Several software programs were developed to solve the linearized or non-linearized form of PBE for diverse shaped molecules, for example Delphi [92, 93], UHBD [94] software.

Poisson–Boltzmann theory is based on the approximation of the solvent as a dielectric continuum and a distribution of counter ions surrounding the biomolecule. The biomolecule is treated explicitly as a collection of point charges centered on the atom, and is represented as a low dielectric cavity (Figure 2-6). The dielectric of the solvent is usually higher than the interior dielectric of a protein due to the degree of freedom of the molecule to move fast and rearrange under the stress of the field generated by the protein. The dielectric reflects the polarity of the solvent. In the protein, ionizable groups are more limited in degree of freedom to rearrange at the surface and inside the cavity. A low dielectric constant for the interior of the protein reflects this limitation and the fact the protein is not highly charged. If significant reorientation of the titrable groups is required, the interior dielectric should be increased [95].

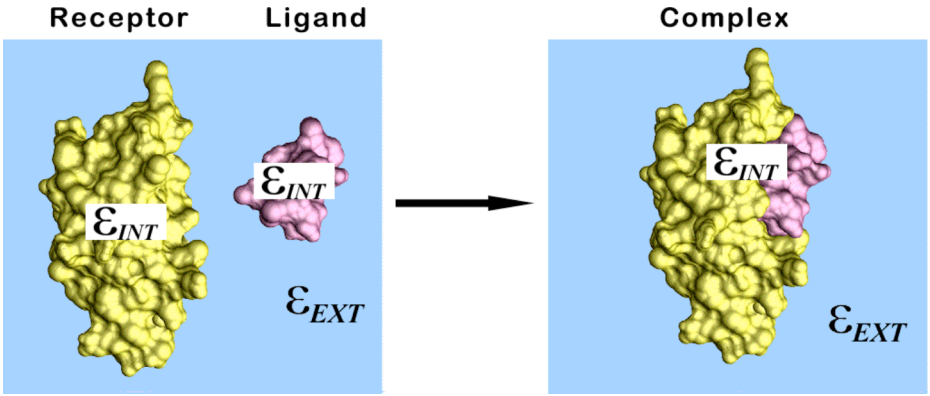


Figure 2-6 Poisson–Boltzmann approach using a high dielectric constant to represent the solvent and a low dielectric for the interior of the biomolecule.

If we consider a continuum solvent polarized locally and linearly under the influence of a generated field, the solvent/solute interactions can be described by Poisson equation:

$$-\nabla.\varepsilon(r)\nabla\phi(r) = \rho(r)$$

Equation 2-11

The Poisson equation defines a dimensionless electrostatic potential ϕ(r) due to the charge distribution ρ(r) in a polarizable continuum with a dielectric constant ε(r). If the system is homogeneous with the same constant ε everywhere, then the equation is reduced to Coulomb's law. Now if the mobile ions are added to the system to count in charge density in the surrounding medium, PBE is finally defined. A boltzmann distribution is used to described the ions distribution in the system.

$$-\nabla.\varepsilon(r)\nabla\phi(r) - \varepsilon(r)\kappa^2(r)\sinh\phi(r) = \frac{4\pi e_c^2}{KT}\sum_i q_i\delta(r - r_i)$$

Equation 2-12

Equation 2-12 is Poisson–Boltzmann equation. The RHS describes the partial atomic charge centered on the atom scaled by a factor (4π) and to ensure dimensionless form of the potential, charge of an electron (e_c) as well as the thermal energy of the system (KT) are included. The LHS is decomposed into a dielectric and electrostatic functions space-dependent, and the second term is related to Debye–Hückel screening parameter to express ion accessibility scale by the dielectric constant of the solvent.

More details about the implementation and details can be found in this review by Sharp [96].

2.4.3 Solvation energy

The interest of the work here is to estimate the interaction occurring between a protein and a ligand upon complexation in solvated environment. During the binding process, the solvent plays an important role as both molecules interact with the solvent surrounding. The energy required to bring the molecules from vacuum to solvent environment corresponds to the solvation energy. Poisson–Boltzmann equation can be solved to estimate the electrostatic contribution to the solvation energy. It is important to estimate the energy of each pieces of the complex when they interact with the solvent. The interactions between the solute and the solvent are estimated with different dielectric constant for inside and outside the biomolecule defined by boundary condition related to the molecular surface. The solute is described as a fixed atomic charge distribution and a mobile charge distribution for the ion in solution is described through a Boltzmann distribution. The solvation energy upon binding can be expressed as described in Equation 2-13.

$$\Delta\Delta G^{solv} = \Delta G^{elec}_{complex} - \Delta G^{elec}_{protein} - \Delta G^{elec}_{ligand}$$

Equation 2-13

Equation 2-13 is the free energy of solvation of protein-ligand upon binding

The right hand side terms are the electrostatic energies for each piece computed in the solvent. The free energy of solvation needs to be expressed as the difference in energy between solvated environment with a high dielectric constant and a surrounding having the same dielectric constant than protein interior. This contribution is called the reference state of the system. So the solvation energy equation can be refined as:

$$\Delta\Delta G^{solv} = \left[G^{elec}_{complex} - G^{ref}_{complex}\right] - \left[G^{elec}_{protein} - G^{ref}_{protein}\right] - \left[G^{elec}_{ligand} - G^{ref}_{ligand}\right]$$
$$\Delta\Delta G^{solv} = \Delta G^{elec}_{complex} - \Delta G^{elec}_{protein} - \Delta G^{elec}_{ligand}$$

Equation 2-14

Equation 2-14 expresses the polar free energy of solvation as a difference in solvation energy between complex, protein and ligand. Each solvation energy is decomposed as a difference in energy between the solute surrounded by solvent with high dielectric and solute immersed in the same environment as the protein interior dielectric.

The energy calculated from Equation 2-14 reflects the change in terms of the interactions of the molecules when they are in a reference state and when they are interacting with the solvent.

2.4.4 Coulomb energy

The intermolecular coulombic interactions in a system result from charge-charge pairwise interactions between all atoms in the molecule. This type of interaction can be relatively strong, for example holding together ionic crystal. They are estimated by applying a uniform dielectric constant between the interior of the biomolecular system and the solvent surrounding. The value of the dielectric is set to the internal dielectric. The calculation of Coulombic electrostatic energy upon binding in a macromolecular complex is expressed as a difference in energy between the complex and the two isolated pieces.

$$\Delta G^{coul} = G^{coul}_{complex} - G^{coul}_{protein} - G^{coul}_{ligand}$$

Equation 2-15

Equation 2-15 represents the coulombic energy contribution to the solvation energy expressed by a difference in coulombic electrostatic energy between complex, protein and ligand in a uniform dielectric equal to the internal dielectric constant of the macromolecule.

2.4.5 Non electrostatic term

Immersing molecules into a solvent generates a nonpolar contribution to the free energy of solvation. The solvent molecules have to reorganize themselves and molecules of the solvent interact with the molecules of the solute **[97]**. First the solvent organization is disrupted to let the molecules coming in and to form a cavity to enclose the solute molecule. The cavitation is related to repulsive non-polar energy estimate from the amount of surface and volume of the solute exposed to the solvent. The exposure of molecular surface to the solvent generates new interactions with the solvent to stabilize the system. Attractive van der Waals interactions occur between solute and solvent molecule and can be expressed through Weeks-Chandler-Anderson integrals-based method.

$$\Delta G^{apolar} = p\Delta V + \gamma \Delta SASA + G^{disp}_{attr}$$

Equation 2-16

Equation 2-16 expresses the apolar energy contribution to the binding energy expresses as a sum of different repulsive and attractive energies involved in the

solvation process. The first term relates the volume accessible of the solute by solvent scaled by the solvent pressure. The second term involved the solvent-accessible surface area of the solute scaled by solvent surface tension coefficient. The third term represent the dispersion energy generated by the new interaction between the residues at the surface of the solute in contact with the solvent.

2.5 Application for the grid

Today grid computing [98] is defined as a dynamic virtual environment. Aims of grid technology are to enable the management and the sharing of high-end resources under a security policy to provide a easy access to data storage and processing power over the network. And the most important thing is that it provides the ability to integrate multiple distributed and independent computing resources. The resources can be spread out on workstation network or on clusters confine within an organization or a public collaboration. The most popular and geographically distributed example of public grid computing effort is the seti@home project [99]. This peer-to-peer computing project takes advantage of the unused CPU time of PCs shared by thousands of people over the world, to scan the universe to find a radio signal proving the existence of an extraterrestrial intelligence. The use of public computational grid becomes more and more used in different fields of science. The bioscience community launches a big project called docking@home [100] to elucidate protein-ligand interaction to help in the discovery of new drugs, as well as folding@home project to study the folding process of proteins.

The grid system gives the capability to intelligently and in a transparent way to select the resources needed for a task asked by a user. An other layer is added to the grid infrastructure with the support of grid workflows that are special scientific workflows [101] involving high performance and/or high throughput computational tasks.

The advances in computational and grid technology offer the flexibility for accommodating computational strategies implying large volume of calculations. Computational chemistry field is resorting to grid computing to perform complex simulation and process large amount of data. The main difficulties in using quantum chemistry codes in the grid arise in the complexity of the software and in running the

application on heterogeneous environments. Most of the programs now are adapted to run on distributed environment like clusters but still effort are done to enable the code to run on different architecture.

The grid architecture includes different layers to hold together collections of resources such as network, computing and storage elements and to provide services such as discovery, monitoring and security for example. Heterogeneous resources compose the hardware layer of the grid architecture with different OS, machine architecture and different storage data system. The computing elements resources are the backend of the system where the application is executing. Storage elements can be local or remote to the computing element, and Network File System or GridFTP server are used to perform the storage. The next layer, called Middleware, corresponds to a class of software technologies that manage the complexity in distributed system and integrates the heterogeneous resources into the grid. The Globus Alliance collaboration released an open source middleware called Globus Toolkit or GT4 (fourth version of GT). GT4 is now used in different grid projects to start the grid infrastructure as it provides tools to manage security, discovery, and monitoring system. Globus is considered as the lower middleware layer and can be coupled with an upper middleware layer to leverage its technology by providing new services. Tools specialized in parametric modeling system like Nimrod [102] for example build new technology on top of the existing Globus. Nimrod is a resource broker specialized in performing parameter sweep application on distributed computing environment. The top layer of the grid infrastructure corresponds to the interaction of the user with the grid infrastructure. End-user tools are part of this last layer and gather user-oriented application like portals or client API. A lot of efforts are put in developing applications to facilitate the use of grid computing. Rich-client interface like framework offer to the user an easy way to discover remote grid applications. The framework GEMSTONE was developing for the scientific community to help researcher in chemistry or biochemistry field to access grid resources and application in a transparent way. It supports also workflow technology providing a powerful tool to perform repetitive tasks.

In the current work Nimrod upper middleware was selected to perform parametric study on the different positioning of the ligand. The combination of the different

methods details in this chapter, is done by integrated the different steps into an experiment. The description of the experiment is set into a plan file and can be launched on the grid infrastructure by the mean of Nimrod.

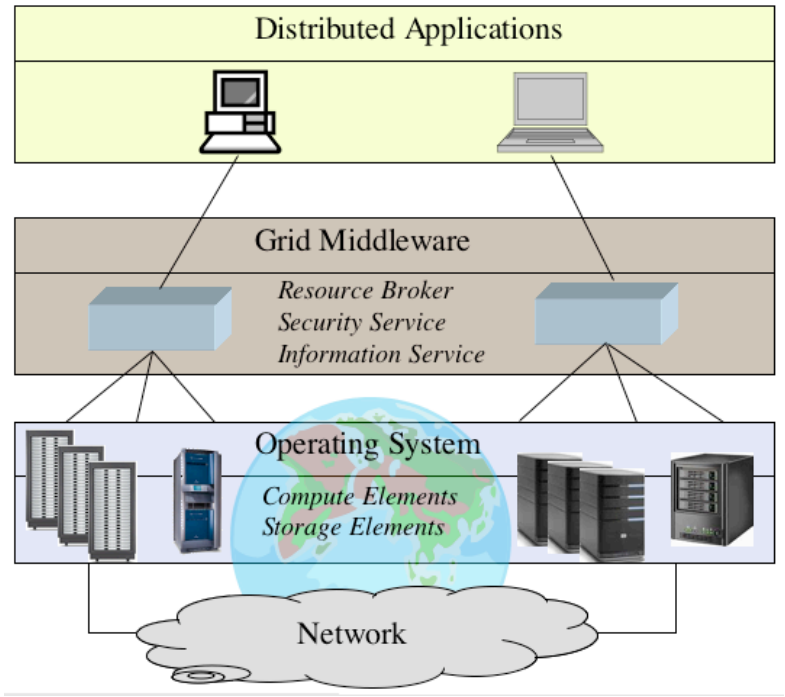


Figure 2-7 Schematic representation of different layers involved in grid architecture.

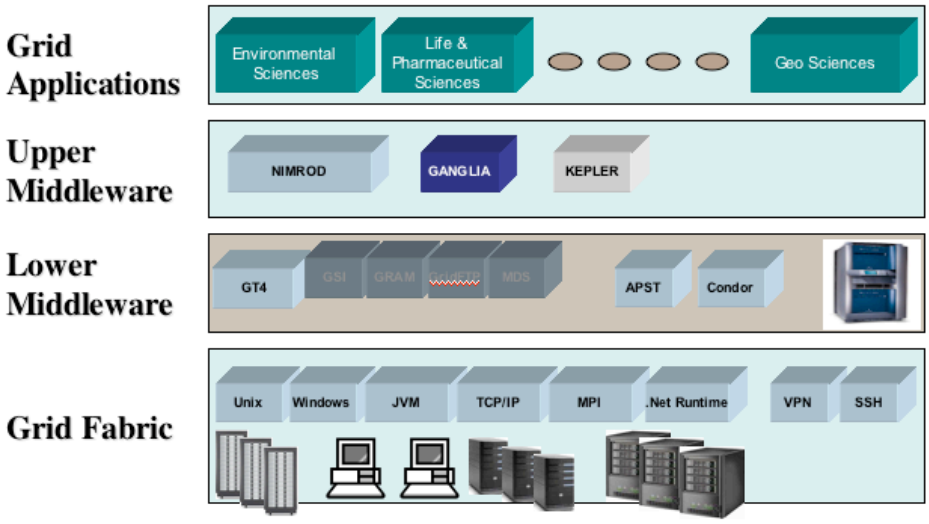


Figure 2-8 Grid infrastructure overview

2.6 Conclusion

Different level of theories are required to perform a docking study on a complex protein–ligand. Force field based methods are used on the protein whereas the small size of the ligand allows a higher level of theory. Quantum mechanical methods are applied on the ligand to get accurate atomic charges and radii to integrate into the electrostatic calculations. Then molecular docking procedures start by screening systematic geometry combinations of the ligand, and poses of the ligand onto the targeted receptor. The combinatorial method gives a large amount of conformational possibilities. Applying geometric constraints screens out the complexes with inter- and/or intramolecular clashes. The method assists the researcher in the quest of finding a structural match between a ligand and its targeted receptor. The ranking is orchestrated by a scoring function designed by integration of different energy components relating the interactions occurring in the binding region.

Technology like the grid computing opens the door to very large-scale studies involving intensive and demanding computations. End-user framework integrating scientific workflow techniques supports the deployment of large docking investigations. They offer a specialized working environment for scientific to automate procedure involving repetitive tasks. The grid infrastructure manages the distribution of computations other heterogeneous resources, helping the researcher in the fastidious task of launching jobs by logging on different machines.

The different methods describe here will be integrated in a hybrid method described in the following chapter.

3 Hybrid method

3.1 Introduction

In the last few years, more and more structure-based drug design research has focused on incorporating more accurate methods, such as offered by QM models, to solve large-scaled biological problems. QM can remarkably improve accuracy for description of molecular interactions with a range of structure and property based descriptors. However, even with the spectacular advances in computational power as well as faster algorithms, computing wavefunction for large proteins remains an enormous challenge. In the present work, we are interested in prediction of structure, energy, and properties of ligands with the potential for high binding affinity to specific sites within macromolecules.

The structure-based model presented here, begins with the preparation of the structure of both ligand and protein. The structural investigation starts by generation of suitable conformational isomers. QM computations are applied on the ligand to get a good insight on the stability by computing its internal energy and providing accurate descriptors as atomic partial charges required for electrostatic study. The electrostatic computations are based on continuum dielectric model to represent the solvent as a continuous medium with ionic concentration. Poisson-Boltzmann Equation is solved to estimate the solvation energy contribution to the binding energy. The interactions of the protein/ligand into solution require the estimation of the energy needed to put the complex into the solvent. This energy is described here as a non-polar contribution. Over the geometry investigation, the energetic of the system based on QM and Classical methods is recomputed with new parameters to predict the new binding energy for the system.

Methods involved in the hybrid model will be described in details in the following chapters to end with a complete function combining the different level of theory to score each conformation of the ligand inside the binding site. In the present hybrid method the definition of scoring function is a slightly different compare to usual term. A scoring function aims to score the different poses or conformations generated during the docking study. Then the ranking is performed relative to the initial structure, which is supposed to be the minimum energy. Here the term scoring

function is used to describe the different energy contributions to the binding energy. The first step in docking process is a structural analysis to eliminate all the conformations with steric clashes. The scoring function is used after this first control, to evaluate the affinity between the two molecules. By summing the energy computed for each type of interactions, the scoring function will drive the results to a set of conformations susceptible to bind the receptor that will become activated or inactivated. More terms can be included into the scoring function as a specific distance between atoms of the binding site and the ligand. The distance will be a constraint to fill to enable the reaction inside the binding site.

The Scoring function (SF) designed for this is:

$$\Delta E_{total} = \Delta G_{coul} + \Delta\Delta G_{polar} + \Delta G_{nonpolar} + \Delta E_{\text{int}} + \Delta G_{conf}$$

Equation 3-1

ΔG_{coul} is the pure coulombic interaction happening in vacuum for the complex, $\Delta\Delta G_{polar}$ stands for the polar solvation contribution to the binding energy computed with APBS, $\Delta G_{nonpolar}$ is an apolar solvation contribution expressed first as a linear function of the molecular surface area and completed by solvent accessible volume term and van der Waals interaction between solute and solvent, ΔE_{int} represents the ranking energy of the ligand estimated by QM methods and ΔG_{conf} is a term quantifying a specific interaction occurring in the binding site between a residue of the receptor and the ligand or atoms of the ligand itself.

Throughout this chapter toolboxes according to the step in the hybrid method will be shown. Then the tools will be assembled to build the flow of sequence to describe the process of predicting binding energy between the ligand and its targeted macromolecule. Figure 3-1 represent the full toolbox with the board on which the workflow will be built.

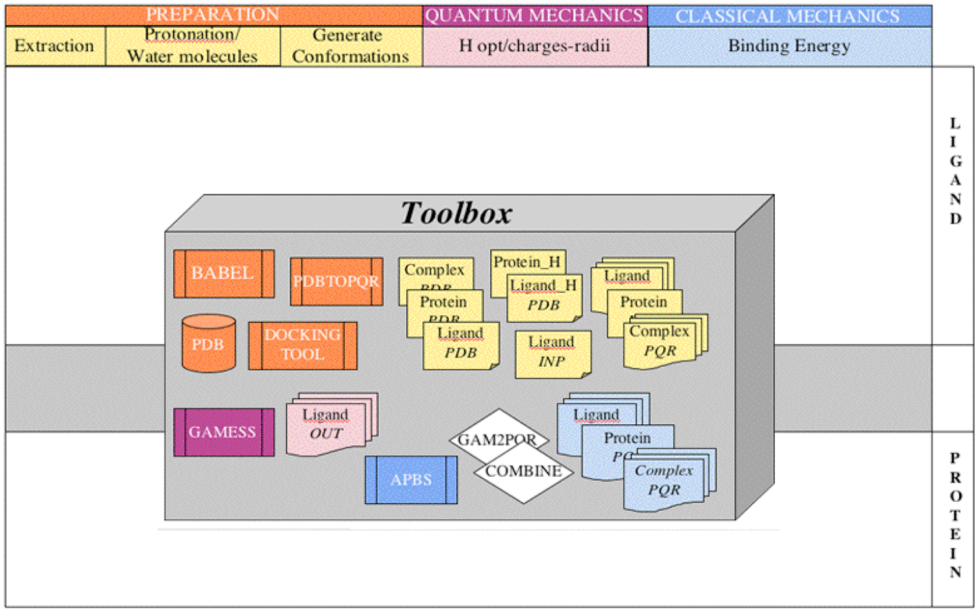


Figure 3-1 Toolbox with all the programs, scripts and files required to integrate the hybrid method into a workflow

3.2 Structure-based method

The discovery and development process of new drugs or compounds is part of rational drug design implying a wide range of computational methods to design and improve the efficiency of drugs. The rational drug design methods can be divided into subcategories as structure-based drug design, ligand-based drug design, de novo design, homology modeling. The present hybrid method uses the former category and so required good structural information of the protein target and the ligand. The method starts with known three-dimensional atomic coordinates of the drug target, most often derived from x-ray diffraction or nuclear magnetic resonance (NMR) techniques. Then high-resolution structures of proteins and potential drugs are stored in repository like PDB for macro-biological systems or Ligand Depot for small molecules. The accuracy of the system's structural details provides a good starting point to study the binding interaction between the macromolecular receptor and the

potential drug compound. In the hybrid method this accurate level is a key to drive a successful analysis on the binding interaction.

The 3D structure of the bio-systems are extracting from the protein databank under their bound form. It is well known now that a protein can adopt a total different shape when it is in presence of a specific ligand. The work of [103] on the interaction DNA-protein illustrates computationally the contribution of conformational adaptation under binding process. The induced fit answer between protein-ligand upon complexation is not treated entirely in this study, as the complex structure considered corresponds to the shape adopted by the protein when it binds the ligand. So the entropic contribution consequence of the structural rearrangement of the protein will not be included in the total energetic estimation of the binding. The flexibility is applied on the ligand but later in the work certain motion will be introduced in the binding site of the protein.

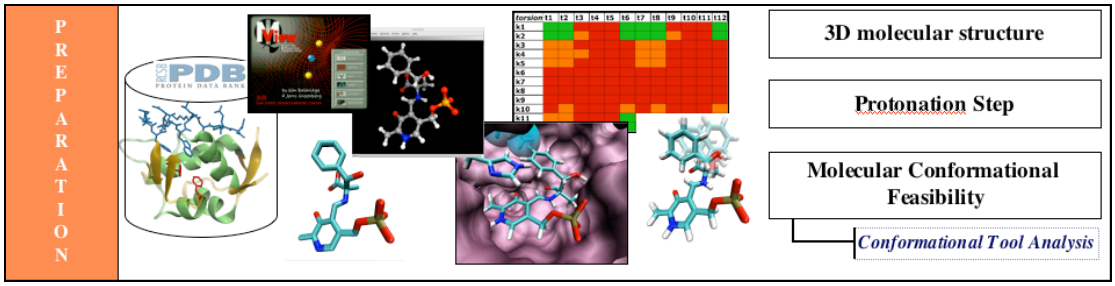


Figure 3-2 Step 1: Preparation steps for the 3D structures

After the extraction of the complex from the databank, the receptor and the ligand are protonated using different programs and investigation on conformations of the ligand is performed through a docking tool. The programs used for this step are listed as part of a toolbox. Along the chapter the toolbox will be empty in order to build the flow of action designed for the hybrid method.

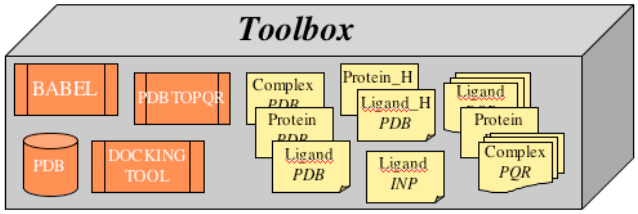


Figure 3-3 Toolbox for the structure preparation step

3.2.1 Structure preparation

The number of 3D macromolecular structure in the Protein Databank exceeds 50 000 entries today. It is an inestimable source of information for researchers. The way in which the structures are deposit and stored are uniform and PDB format file is a standard in the representation and description of biological molecules. The files contains Cartesian coordinates for each atoms classified by amino acids for each chain of the protein. The pdb file extracted from the Brookhaven databank, provides the structure of the receptor with its ligand when they bound to each other and sometimes water molecules and ions are presents due to the crystallographic process. Water molecules need to be removed to treat the solvent implicitly. The ligand and protein coordinates are extracted from the complex pdb file, and two new pdb files are created.

3.2.1.1 Protonation step

Both structures need to be protonated as x-ray diffraction can't elucidate the position of light atoms. The hydrogen atoms are added on both ligand and protein using the chemical software toolbox called Open Babel [104]. The network of hydrogen-bonds and the position of hydrogen atoms in the protein need to be optimized due to the impact on the electrostatic calculations [52, 105-107]. In some structures the imidazole ring of Histidine residue has to be flipped of 180° to recreate local hydrogen-bond interactions, as well as the side chain of arginine or glutamine. So the

protonation state needed to be determined for those residues presented different isomeric forms [108]. The improvement of the hydrogen bonding and the positioning of light and heavy atoms are done via Monte Carlo with a tool called PDB2PQR [109]. The protonation state of the ligand needs to be controled to ensure a conservation of the interactions occurring in the binding site as well as the conservation of the overall charge of the molecule. This checking point is realized by loading the ligand structure into a molecular visualization program, like Sirius [110]. This visualization framework provides features to add or remove hydrogen atoms as well as to modify the hybridization state of atoms, which is convenient to correct the protonation state of molecules.

3.2.1.2 Atomic partial charges and radii for the protein

PDB2PQR package used to prepare the protein structure automatically assigns atomic parameters required for electrostatic study in continuum solvent. The parameters are stored in force field. PARSE [52] force field is selected for the hybrid method because of its parameters optimized for Finite Difference Poisson-Boltzmann calculations. The atomic partial charges and radii were derived from a database of experimental solvation free energy. The choice of charges scheme is important in electrostatic solvation energy determination as this energy depends strongly on the arrangement of the partial charges and their exposure to the solvent [111]. Force fields are based on the topology of the amino acids. Sometimes crystallographic structures stored in the PDB have “un-natural” amino acids. For those types of residues an assignment by hand is required. Following the published data on the radii and the charges from PARSE[52], all the parameters for the non-regular amino acid can be defined. The program PDB2PQR takes in input a pdb file format and builds a pqr file from the structural information of the pdb file and the parameterization of the force field chosen.

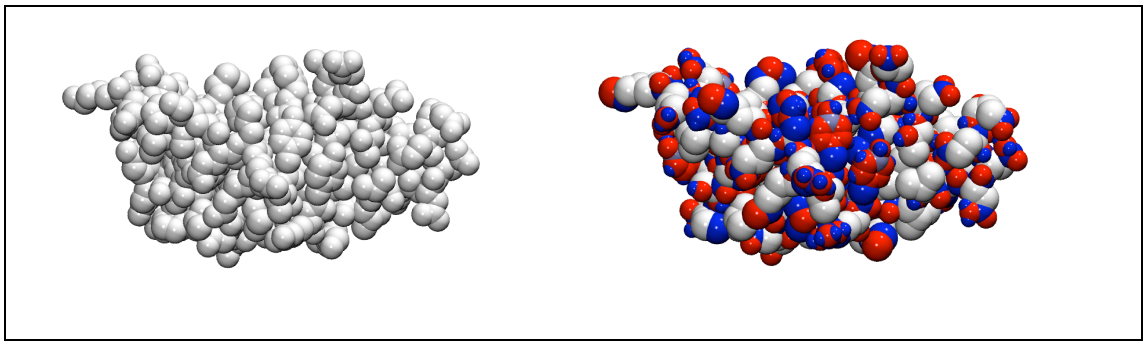


Figure 3-4 Atomic partial charges mapped onto van der Waals radii

Figure 3-4 is the 3D representation of a protein structure with atomic radii modeled and the charge option as color. On the left the pdb file load with no charge and van der Waals radii, on the right the pqr file with the radii and charges assigned by the program pdb2pqr

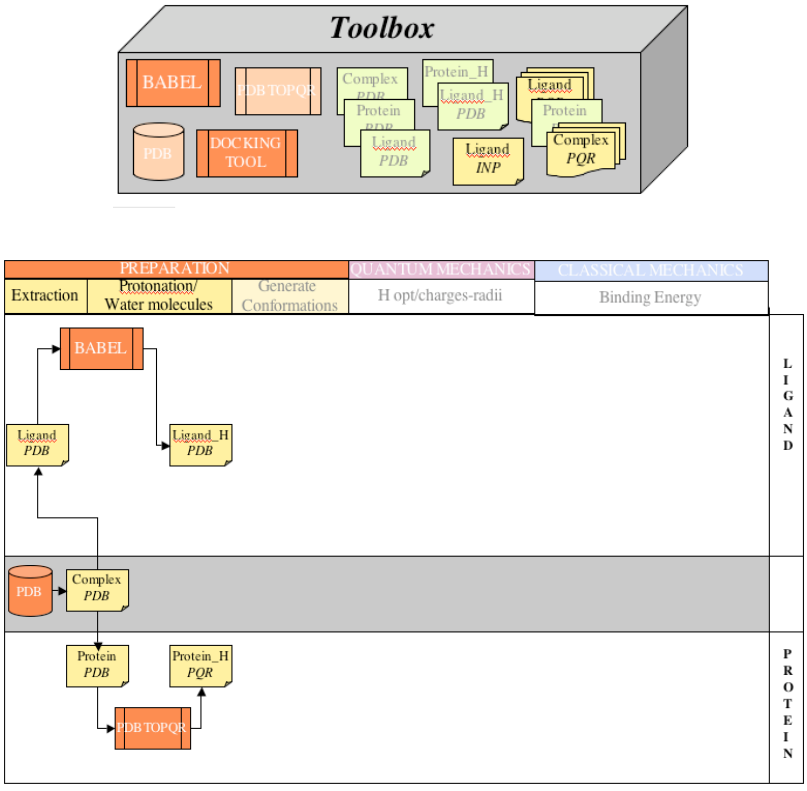


Figure 3-5 Toolbox and workflow state after protonation step

3.2.2 Conformation Tool Analysis (CTA)

The docking investigation, as mentioned in the introduction of this chapter, involves moving the ligand and modifying its conformation to find the best complementarity with the receptor. This analysis is based on sterical properties to find the conformations of the ligand matching the receptor. Each configuration and conformation are analyzed in term of van der Waals overlaps between all the atoms of the system. From this analysis only the good structures will be investigated on molecular recognition to study the non-covalent interactions contribution, which save computation time for non-feasible structures.

3.2.2.1 Conformation generation

A program developed internally in the group enables one to generate new conformations from a starting structure by defining torsion angles. The search begins in a systematic way. The user defines the dihedral angle along which the rotation will be done and the number of steps. The step number corresponds to the angle space between each new conformation. The dihedrals have to be defined by picking four atoms involved in the torsion. This choice causes the split of the molecule into two different pieces. A part remains static along the rotation and is called the static part. The second part corresponds to the atoms moving along the rotation and it is called the flexible part (Figure 3-6). All the conformations of the flexible part are merged with the static part creating a new structural arrangement of the ligand.

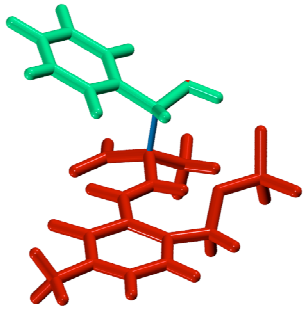


Figure 3-6 Flexibility on one torsion angle of the ligand.

The molecular part colored in red represents the static part of the molecule, the part in green is the flexible part involved in the rotation, and the bond colored in blue represent the anchor of the two part and the rotatable bond of the dihedral angle.

The problem becomes more complex by increasing the degree of flexibility on the molecule. Geometries along several dihedrals can be investigated. A combinatorial study is performed to screen the library of the flexible fragments. The new conformations of each flexible part are merged together. And the overall structure ligand is rebuilt by adding the static part.

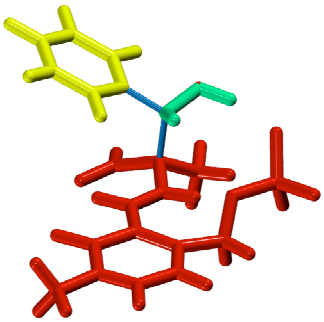


Figure 3-7 Flexibility on two torsion angle of the ligand

The molecular part colored in red represents the static part of the molecule, the part in green is one of the flexible part involved in the rotation, the yellow part is the second flexible part, and the bond colored in blue represent the anchor between the static and flexible part.

The number of steps increases also the complexity of the combinatorial study. Small steps will generate a large amount of fragments. A large space between number of steps might miss good positioning of the ligand inside the binding step. The strategy is to use a middle step number in order to detect which space offers room to fit the flexible part. A refined study can be done again with smaller steps in the range of angles enable the positioning of the flexible part.

3.2.2.2 Conformation checking point

The conformation generation mentioned before, exposes the issue of the rapid increase in complexity by the multitude of new conformations created. The systematic way to generate new conformations can be supplemented by a non-systematic approach. The non-systematic method here means to avoid taking in account all the possible conformations generated by the rotation along the rotatable bonds. After merging all the parts involved in the molecular structure (static and flexible pieces), the distances between the atoms of the flexible and between the atoms of the static part are estimated. The distances vary along the motions applied on the molecule and if two atoms become too close to each other, the repulsion become too strong and the stability of the molecule is not longer preserved. Checking the distances between each atom from the different parts avoids the study of structures having intramolecular clashes. Then molecule is replaced in the complex, if the flexibility was applied on the ligand, the new conformation of the ligand is putting back into the binding site. The distances between the atoms of the new conformation of the ligand and the atoms of the receptor has to be checked. A control is applied to sort out the complex formed with atoms of the ligand too close to atoms of the receptor. So the conformations of the new complex having intermolecular clashes are eliminated. More flexibility is allowed on this checking point, as the system in vivo is not static. We consider that having a bigger amount of overlap between the van der Waals radii of the atoms of both species will be more realistic.

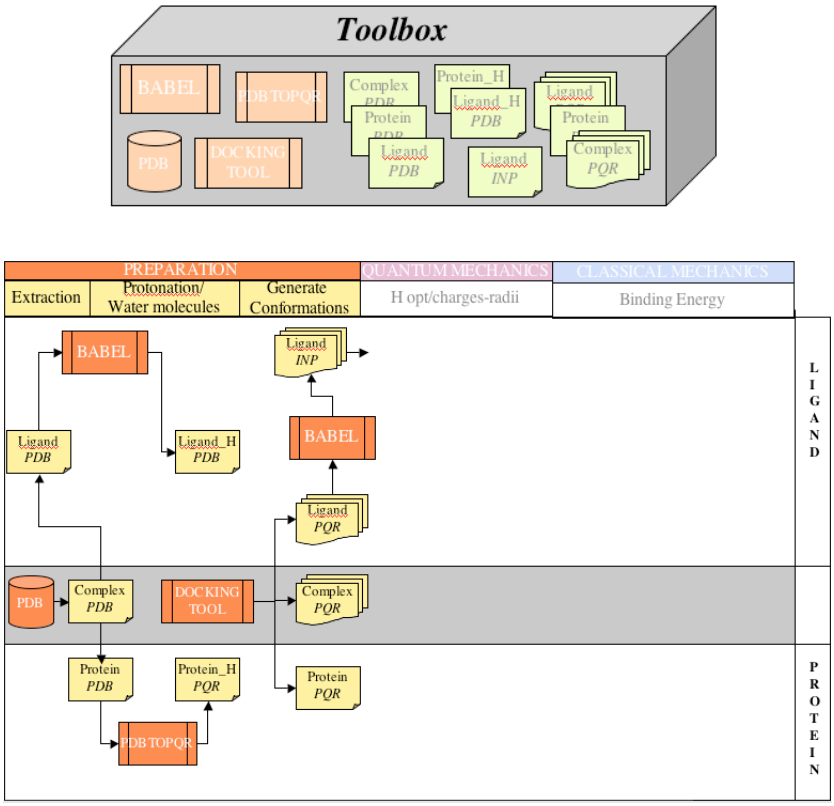


Figure 3-8 toolbox and workflow state after the entire preparation step

3.3 QM investigation

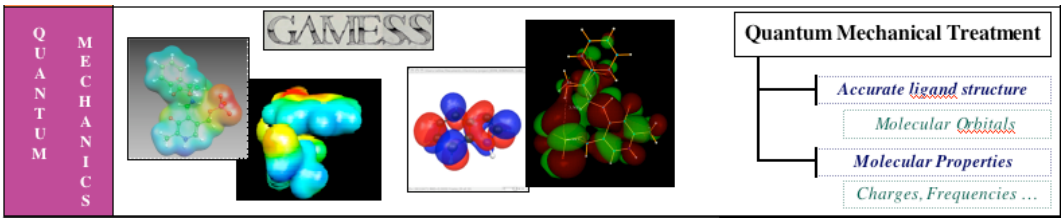


Figure 3-9 Quantum Mechanical treatment of the ligand with GAMESS to get accurate properties.

3.3.1 Optimization of Hydrogen atoms position

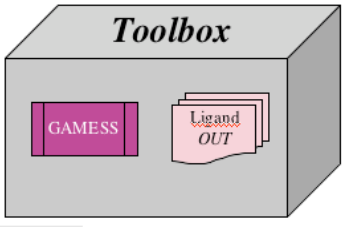


Figure 3-10 Toolbox for QM investigation on the ligand

The program Babel has the feature to assign hydrogen positions onto the ligand structure. The hydrogen addition is based on empirical methods and their positions need to be optimized. *Ab initio* quantum chemistry methods are applied to refine the positioning of the hydrogens atoms. The computations have been carried out using the program GAMESS [55]. The calculations are performed at B3LYP[112-114] level of theory with a split-valence double-zeta basis set. Polarization functions are added to the as p and d basis functions, to increase the flexibility of the basis set. Diffuse functions are added on the hydrogen atoms and the heavy atoms to improve the convergence to basis set limit. And allow a better diffusion of the partial charges.

If the study involved a conservation of the main structure of the ligand, constraints need to be applied. Providing the list of Cartesian coordinates numbering can freeze the heavy atoms. The hydrogen atoms are the only one allowed to move along the iterations. Parts of the ligand can be fixed if the study requires it and only a portion of the molecule will undergo the electronic rearrangement with optimization of the placement of the atoms along the self-consistent reaction field. The minimal energy after optimization convergence represents the internal energy of the system (E_{int}). This internal energy reflects the stability of the structure and it is an important criterion to include in the equation to score the different conformations.

3.3.2 Partial atomic charges

Electrostatics driven by the Poisson–Boltzmann equation requires atomic descriptors. One important parameter is the partial atomic charge. The charges can be derived from force field for big system. A more accurate way to determine the charge is to use charges model implemented at quantum theory level. Such a high degree of accuracy can be only applied on small system. In protein-ligand interactions study, the ligand is usually small enough to be treated fully by QM methods. In the strategy exposed here, the determination of the partial atomic charges of the ligand will be performed with GAMESS. The charges schema used is the one proposed by Brenneman, ChElPG model [115]. Details regarding ChElPG are given in the methodology; this charge model assignment is accurate and adapted for electrostatic study, as it tends to fit the electrostatic potential.

After running Gamess, the charges need to be extracted from the output file. A fortran program written in the group is parsing the output file to extract the charges computed with ChElPG schema. In the same time the program is extracting the optimized coordinates of the structure resulting from hydrogen optimization positioning or more optimization part of the molecule. The data extracted are stored in a file format with pqr extension. It is similar as pdb format except that the atomic charges and radii are replacing the last two columns corresponding to temperature and occupancy. The atomic radii are derived from PARSE force field chosen also to treat the protein as the parameters are improved to rely on FDPB. This new file format will be used for the electrostatic calculations to provide the parameters required to solve PBE.

3.3.3 Internal energy of the ligand

The geometry analysis of the ligand produces a lot of conformations. A program controls the feasibility of each structure to check if intramolecular clashes occur along the rotation on flexible parts of the ligand. This checking point is really important as it eliminates wasting computational time especially at the most time consuming level. The stability of the molecule depends also on physico-chemical properties. A way to estimate the degree of stability of a small molecule is to compute its internal energy. In drug discovery the stability of the potential candidate is a key success. The internal

energy reflects the stability in term of physical properties as bond length or angles but also in term of electronic arrangement.

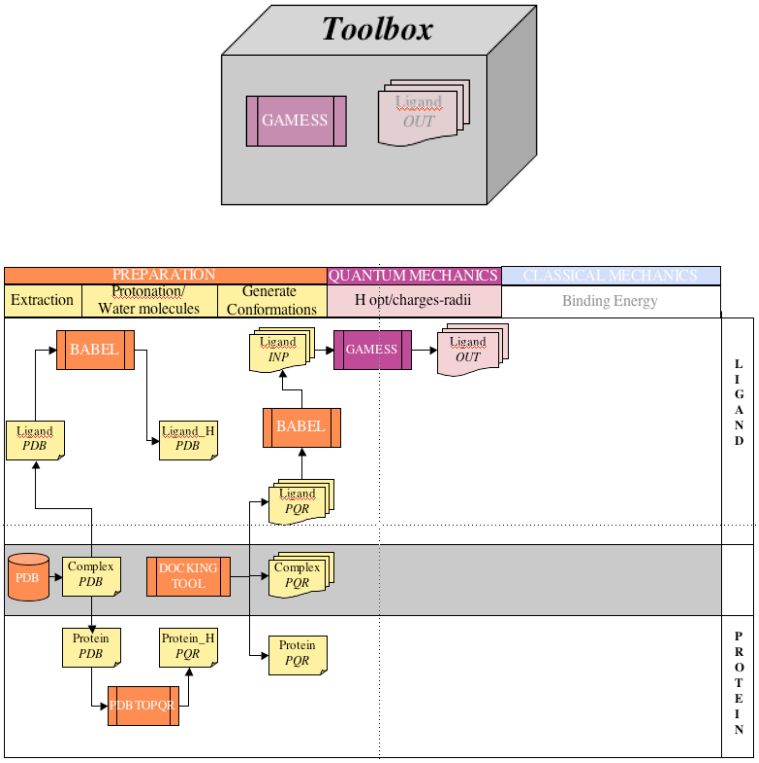
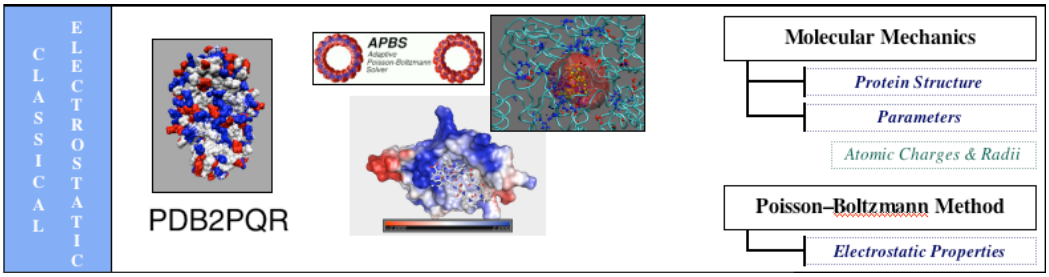


Figure 3-11 Toolbox and workflow state after QM treatment on the ligand

The surrounding environment of the ligand when it is embedded in the pocket of the protein, need to be reproduced in QM computations. Tomasi and coworkers implemented a Polarization Continuum model [69] (PCM) in Gamess. This solvent model represents the solvent as a dielectric continuum characterized by a dielectric constant. This method can be applied to compute the internal energy of the ligand with the influence of the field generated by the protein. The interior of the protein is represented as a low dielectric constant of 4 [116]. Computation carried in solution with Gamess will be done at a dielectric 4 equivalent to chloroform solvent.

3.4 Electrostatic Investigation



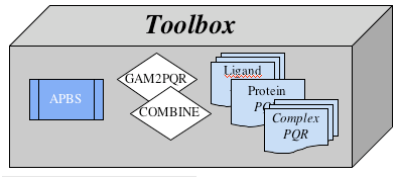


Figure 3-12 Toolbox for electrostatic investigation

Many methods have been developed to study electrostatic interaction occurring during a binding process between a protein and a ligand. In the method section we have discussed the use of Poisson–Boltzmann (PB) theory in the situation of solvent containing mobile ions. This dielectric continuum model is fast and accurate in determining the electrostatic contribution to the free solvation energy of a biomolecular complex [117]. Today enormous progress has been achieved in Generalized Born [15, 21] method that reach a good level of accuracy and in a faster way than PBE. PB theory stands the most accurate and widely applied in biological and chemical science. PBE for complex system need to be solved numerically due to complex charge distribution and challenging surfaces shape. The analytical solution of PB can be applied only for cases with simple geometry. The work of Professor Nathan Baker has improved the efficiency and accuracy of the algorithm [118, 119]. Simplification are usually made to reach the linearized form by truncating the Boltzmann factor (first-order term in Taylor series) in the expression of the ion distribution. Moreover, with huge improvements in computing power, PBE becomes

accessible to estimate electrostatic interaction between a protein and its ligand in a continuum salty solvent. Solving PBE is fast enough to combine it with molecular mechanics method to estimate the binding energy for each poses generated [120, 121]. Software like DelPhi [92, 93], UHBD [94] or APBS [122] are available to perform PB calculations. The software package use in this study is Adaptive Poisson-Boltzmann Solver (APBS). It solves numerically Poisson-Boltzamnn Equation (PBE) by means of finite difference/finite element approach.

3.4.1 Solvation energy contribution

The strategy here is to use APBS to estimate the binding energy between a protein and a ligand in solution. The free energy of binding is decomposed in several contributions. The first contribution is the solvation contribution obtain by subtracting the electrostatic energy of each isolated molecule to the complex, as shown in Equation 3-2

$$\Delta G_{solv} = G^{elec}_{complex} - (G^{elec}_{protein} + G^{elec}_{ligand})$$

Equation 3-2 Solvation energy contribution to the binding

Equation 3-2 expresses the solvation contribution to the free energy of binding estimated by solving the non-linear partial differential equation of Poisson–Boltzmann by using an adaptive multilevel finite element method with APBS.

APBS carries out the computations on a parameterized grid. The three-dimensional grid size has to be large enough to capture the whole complex. It is important to define the grid with the same set of parameters for the protein, ligand and the complex to delete the self-interaction term. A high dielectric constant is assigned for the solvent; water has a solvent dielectric around 80. A low dielectric represents the protein interior, typically around 4.

Another contribution to the free energy of solvation has to be taken in account. It is the energy of each molecule in a reference state. As explained in the method chapter this contribution energy corresponds to apply coulombs law with a scale factor equal to the interior dielectric. Subtracting the reference state energy to the electric energy

of each molecules reflect the energy require to move the molecules from gas phase to solvent surrounding.

$$\Delta\Delta G^{solv} = \left[G^{elec}_{complex} - G^{ref}_{complex}\right] - \left[G^{elec}_{protein} - G^{ref}_{protein}\right] - \left[G^{elec}_{ligand} - G^{ref}_{ligand}\right]$$
$$\Delta\Delta G^{solv} = \Delta G^{elec}_{complex} - \Delta G^{elec}_{protein} - \Delta G^{elec}_{ligand}$$

Equation 3-3

3.4.2 Apolar contribution

3.4.2.1 Apolar energy strategy I

There is an apolar contribution to the free energy of solvation. It results mainly of the work required to create the cavity in the solvent. It is proportional to the difference in Solvent-accessible surface area [123-125] between the complex, protein and ligand. This method to estimate the non-electrostatic contribution was used at the beginning of the project. The subtraction of the solvent-accessible surface area (ΔSASA) represents the surface area of the protein and ligand that lost or gain the contact with the solvent; it is the buried or exposed surface. This ΔSASA term is then weighted by the apolar surface tension γ.

$$\Delta G_{nonpolar} = \gamma \Delta SASA$$

Equation 3-4 apolar contribution term proportional to the Solvent-Accessible surface Area

3.4.2.2 Apolar energy strategy II

A second strategy was derived during this thesis. Poisson–Boltzmann remains a good theory to apply on dilute protein systems to reproduce the polar explicit solvent forces but the popular solvent-accessible surface area model is not accurate enough to describe non-polar solute-solvent interactions. As reported by Nathan Baker *et al*, in his paper published in 2006 [126], the scaled particle theory(SPT)[127, 128] from which the SASA model is derived, described poorly the energy required to create the cavity. The Solvent-Accesible Volume (SAV) quantity needs to be added. So SASA and SAV terms expressed the repulsive part of the nonpolar solvation interaction. The attractive solvent/solute nonpolar solvation interactions are described as van der Waals interactions between solvent and solute[129]. This dispersion terms can be decomposed as Weeks-Chandler-Andersen (WCA)-like integrals[130, 131]

$$\Delta G_{nonpolar} = E_{rep} + E_{attr}$$
$$E_{rep} = \gamma SASA + pSAV$$
$$E_{attr} = E_{disp}$$

Equation 3-5: nonpolar contribution to the free energy of solvation

3.4.3 Coulombic Contribution

The Coulomb energy is calculated with simple Coulomb's law in a uniform dielectric constant turn at 1. This energy relates the pair wise interactions between two partial atomic charges separated by a certain distance in a dielectric of 1.

$$\Delta G_{coul} = \Delta G^{coul}_{complex} - \Delta G^{coul}_{protein} - \Delta G^{coul}_{ligand}$$

Equation 3-6 Contribution of coulombic energy to solvation energy

3.4.4 Parameter to solve PBE

3.4.4.1 Atomic charges and radii

Poisson–Boltzmann equation required atomic charges in order to estimate the electrostatic potential. Those charges are defined accurately by QM method using CHELPG models as described in the QM section. To extract the charges from the ouput file of GAMESS, a program was written to generate the pqr file of the ligand used by APBS.

3.4.4.2 Dielectric constant for continuum electrostatic study

In continuum methods the dielectric properties of the solute, the protein, and the solvent are represented by a dielectric constant. Whereas the solvent dielectric constants are well known and reflect their polarizability, the interior dielectric constant [96, 132] of protein is more difficult to quantify. This difficulty comes from the complex organization of the protein with buried and exposed titrable groups that perturb locally the reorganization of permanent dipoles and modify the interaction in the surrounding. This is an active subject of discussion in the scientific community. The measure of dielectric of protein in a dry environment is usually 2-4 according to experimental results[73, 74]. According to the theoretical study of Nakamura et al [133] on dielectric constant of protein the value will fluctuate between 1 to 20 whereas 2 to 4 for Honig and Gilson [116]. Actually there is a big discussion on higher dielectric constant[134] even the use of homogeneous dielectric model (dielectric of the solvent applied in the interior of the protein) had been demonstrated to give good results compare to experiments. The choice of the dielectric depends also on the continuum model used to study a particular macromolecular system. Warshel and Schutz[95] study different implicit models and the way to determine dielectric constant value adapted for those type of electrostatic models. They come to a conclusion that using a model based on Poisson–Boltzmann required a dielectric constant between 4-8 but it seems that a value of dielectric 4 is more likely to be a correct estimation.

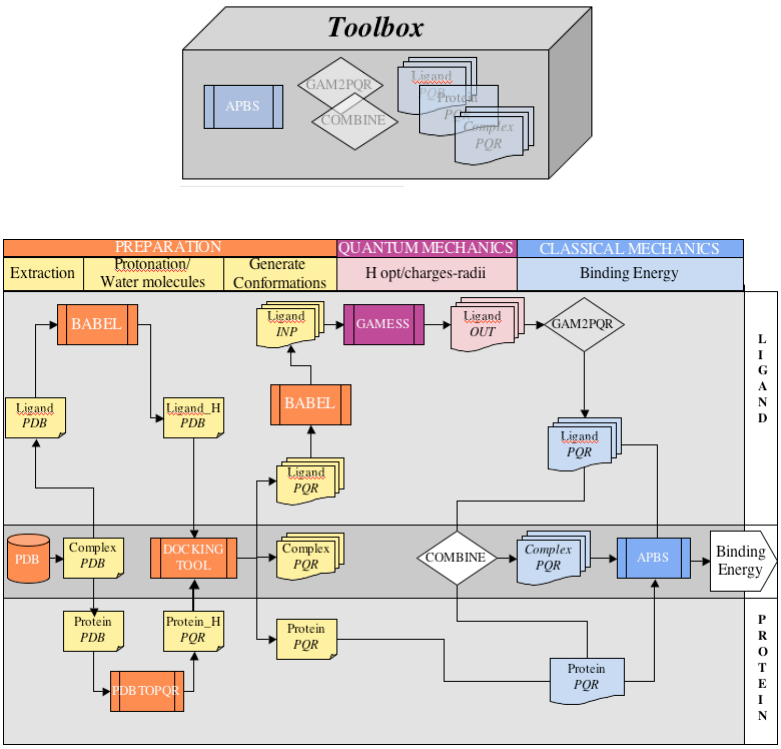


Figure 3-13 Workflow encapsulated all the steps required for the hybrid method

As we mentioned in the previous chapter the difference in entropy upon binding will not be taken in account as a reflection of the structural adaptation of the receptor and the ligand to bind together. The structure upon complexation provides the Cartesian coordinates of both molecules in the position they adopted when they are put together in a salt-aqueous environment. In order to improve the accuracy on the best poses for the ligand to bind the protein, the conformational energy change of the ligand will be including in the overall energetic. From this entropic term will be estimated the stability of the small molecule as a criterion of feasible conformation. Applying too many constraints can generate unstable molecule if there is no compensation from the surrounding.

3.5 Scoring function

Score $\Delta E_{affinity} = \Delta E_{steric} + \Delta\Delta G_{elec} + \Delta G_{non-elec} + \Delta E_{QM}$ **Scoring Function**

The scoring function gather together all energetic components calculated along the study. The conformations generated with the CTA will be investigated following the steps describe above. First an internal energy of the ligand will be estimated by QM methods using GAMESS. The solvation contribution to the binding energy will be calculated with APBS with corresponding dielectric parameters. Accurate atomic charges computed with QM method for the ligand will be integrated into PBE by generation of a PQR file as an input file for APBS. The coulombic term will be estimated with APBS by setting the dielectric constants to 1 in order to estimate the energy in a vacuum environment. Finally the non-polar contribution to the binding energy is estimated with APBS using a flat file containing a set of empirical parameters needed to solve the WCA integrals to estimate van der Waals interactions.

$$\Delta E_{total} = \Delta\Delta G_{polar} + \Delta G_{nonpolar} + \Delta G_{coul} + \Delta E_{\text{int}}$$

Equation 3-7 Scoring function designed for the hybrid method to rank the different complex conformations

3.6 Control through the grid

The first step of descriptor files of the molecule is using a molecular modelization program to build the molecule. Then the file stored on the local workstation is copy to a remote resource via secure protocol over the network. The resources are usually clusters on which applications like quantum chemistry code run more efficiently due to the parallel distributed computing environment. The researcher launch several jobs required for his study and then need to copy the resulting files to its local machine for analysis. All the tasks cited above typically require tedious interactions with a command line interface via remote connections the user has to manage. In order to provide an easy way to deal with remote resources and coordinate the use of several programs, rich client applications are developed to offer a user-friendly interface to the researcher to integrate all the features required into portals or framework. These components as mentioned in the methodology section are built on top of the middleware layer. A middleware application provides transparent access to remote machine to the user though a desktop application. The application can also combine a user interface for tools and programs used for a specific domain of research. When the tasks are repetitive and can be link together in a logical way, workflow technology can be integrated into the client application to facilitate the work of the researcher.

3.6.1 Workflow design for the hybrid method

The hybrid method required the use of different tools and programs to investigate the binding affinity between a small molecule and its target receptor. The different tools inherent to the method need to be linked by extracting data from output files generated by software to create input files for the next software. Tools have been developed to bind the different applications and scripts are written to facilitate the submission of large number of jobs. A detailed and coherent workflow needs to be designed to integrate the series of tasks. The power of the grid can be combine with the logical sequential actions of the workflow. As shown on Figure 3-14, the hybrid method is decomposed in three big steps.

The first step consists in preparing the ligand and protein structure extracted from the PDB for the study. Hydrogen atoms are added to each molecule. Babel program is used to add hydrogen atoms to the ligand pdb file extracted from the Brookhaven database. PDB2PQR program is used to add the hydrogen atoms on the protein according to the force field selected. The program take in input the protein pdb file extracted and output a pqr format containing the atomic charges and radii defined by the force field. The conformational analysis carried on the ligand is performed with programs written in our group. These programs provide feature to perform rotation along torsion angle by specified the four atoms involved in the torsion and the step between each rotation. Translation and rotation motion can be applied on the ligand to investigate different poses. A program checks the steric fitting between the ligand and the protein and also the distance between the atoms of the ligand when torsion motions are applied on it. A minimal distance can be defined to sort out the feasible conformations from the structure showing clashes between atoms of the two different parts. This minimal distance can be defined as a distance inferior to the sum of the van der Waals radii or the sum of the atomic radii. A percentage of overlap can be tolerated by using another option of the program, which is more realistic as the system is in movement and not static. The programs generate pqr files with the new coordinates of the molecules, the number of files generated depends on the size and number of steps enter for the rotation along torsion angle or the number of translation and rotation ask for.

The second step is the QM treatment of the ligand. Using the software Gamess requires a special format for the input file; so the pqr file of the ligand from step 1 need to be converted into a INP format. The conversation step can be done with BABEL by giving the pqr file in input (similar to a pdb format). The input file generated has a minimal number of keywords. Dollar groups and keywords required for the run need to be added. A script PQR2INP provide this conversion step and in addition write the list of heavy atoms that needs to be frozen. Computations are carried out to optimize the hydrogen position and get accurate parameters for the electrostatic calculations.

The last step is continuum electrostatic investigation based on Poisson–Boltzmann methods. The free energy of solvation is estimated by running APBS on the complex.

APBS requires an input files containing all the parameters need to solve the PBE. For the study of complex protein–ligand, each structure is described in pqr format with atomic charges and radii. All the other parameters to solve the PBE are given to the program in an input file format with the extension ".in". The pqr file for the ligand is generated with the program gam2pqr by extracting the charges from the output file of GAMESS and using the van der Waals radii of the atoms [135]. The electrostatic contribution to the solvation energy is calculated. The apolar contribution to the solvation energy is also calculated with APBS by providing a parameter file with atomic charges, van der Waals radii and van der Waals well depth for each atom of residues. The structural description of the molecules is given in pqr format with the Cartesian coordinates. The electrostatic energy and the non-electrostatic energy are added to estimate the solvation contribution to the free binding energy of the complex. A coulombic contribution to the binding energy is added by performing the same computation as for the electrostatic energy with the only difference of dielectric constant that are turn to 1 to represent the vacuum environment. The sum of the different energy term correspond to the binding energy between the ligand and its receptor.

The results of each step are integrated into the scoring function to rank the different conformations and poses of the ligand related to its affinity to bind the biomolecule. Investigation were carried out also with flexibility apply on the receptor side by rotating ring of aromatic residues.

An automatic procedure such as this allows expanding of the investigation in the conformational fitting between the receptor and the ligand, and performing more accurate computations to understand the nature of this binding by estimated the binding energy of the complex.

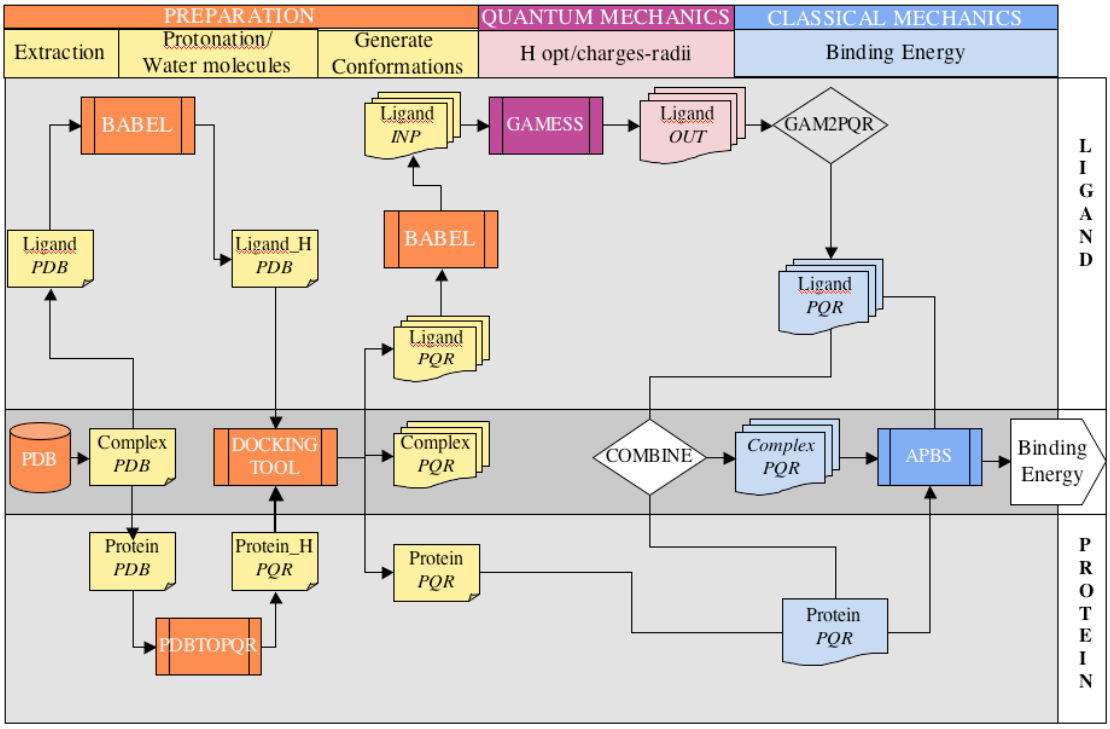


Figure 3-14 Workflow design for the hybrid method.

3.6.2 Gemstone framework

A project involving people from SDSC and the university of Zürich was lead by Professor K. Baldridge to develop a framework called GEMSTONE [43] (Grid-Enabled Molecular Science Through Online Networked Environments). It offers a rich client desktop application to access grid resources and scientific applications to use over the grid infrastructure. It supports also workflow engine to let the user building is own scientific workflow for a specific research topic.

Gemstone was developed to complement the hybrid method to combine in an automatic and logical way the different steps inherent to the method. The framework provides a workflow environment to automate the steps described in this chapter. Special tools like MolPrep were created to investigate the conformational space of the ligand. This application was fully integrated in the interface and the user can interactively pick the atoms involved in the torsion and define the parameters for the rotation. The tedious preparation of the molecules can be carried out through Gemstone with a workflow designed to support the hybrid method.

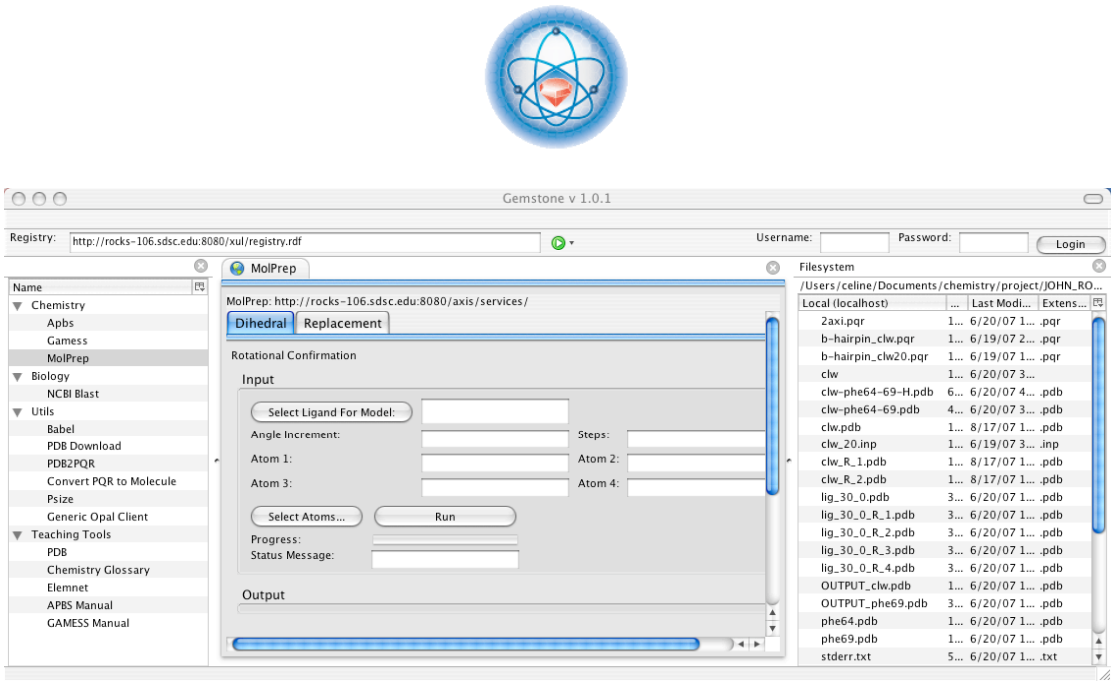


Figure 3-15 Gemstone user interface screenshot.

The interface is decomposed in three main parts: a service registry panel on the left for the scientific applications available, a file system on the right to access on local directories and files, and the middle area contains the interface for the application selected by the user. The MolPrep interface is loaded on the screenshot shown on the Figure 3-15. The user can define the parameters to perform rotation on a molecule by defining the number of steps and the size of step. The interface provides a graphical modeling tool to load the molecule studied and picks the atoms involved in the torsion directly on the 3D model. Then when the user executes the program, the execution is done on remote machine specified by the user with the choice of the registry file.

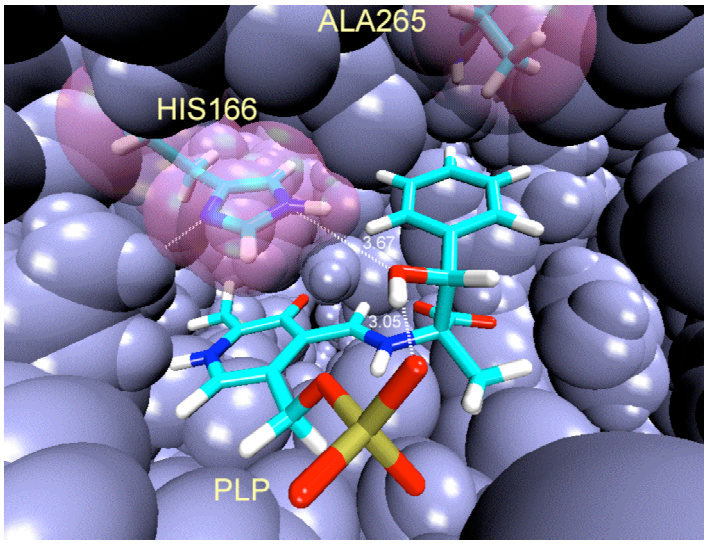
Several applications are wrapped up into the web services to be integrated in Gemstone framework. All the tools used in the hybrid method are available into Gemstone. A tool called Opal was released in the same time as the development of Gemstone and provides the feature to wrap other scientific applications as web services. This tool gives the opportunity to turn in web services the scripts written to create the link between the different tool used along the hybrid method.

3.6.3 Distributed parametric study on molecular conformation with Nimrod: docking tool

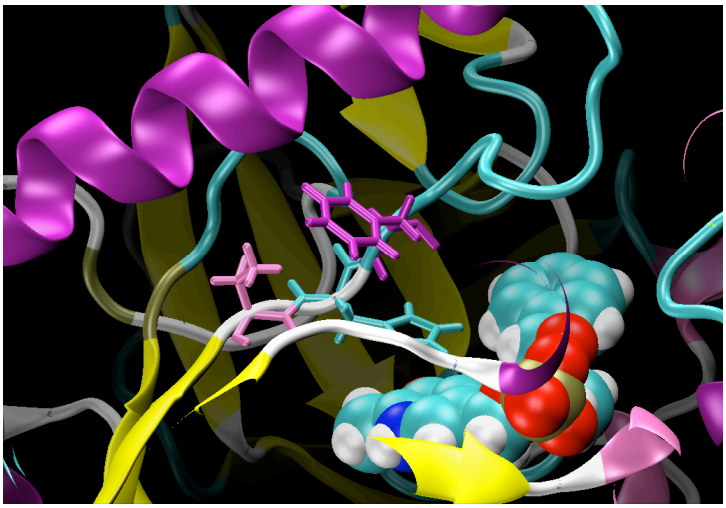
Investigations of the positioning and geometry of the ligand generate many possibilities that require a combinatorial analysis. The numbers of degree of freedom apply on the ligand for example, determines the complexity of the analysis. The number of torsion angle and the size of the step allowed on the rotation generate a certain number of new conformation [136]. For each conformation, the computations have to be carried out to integrate each result into the scoring function. The parameters given to solve Poisson–Boltzmann have to be updated. The new Cartesian coordinates defining the position of the atoms are integrated into the charge distribution expression. The partial atomic charges change with the conformation and the new values are incorporated into the fixed charge distribution of the system. Conformational changes drive to a modification in the different energy components of the scoring function.

The hybrid method determines the affinity between a receptor and a ligand by ranking the different conformations/poses of the ligand along the motion apply on the structure. The rotatable bonds can be defined and a set of conformations can be generated. In this conformational analysis the setting for the rotation along the torsion can be integrated as parameters in a parametric modeling application. The middleware Nimrod/G has the capability to perform parametric computations on geographically distributed resources. The services include the management of the resources to schedule the computations. In the present work, Nimrod /G give the opportunity to drive a parametric study based on the torsions setting by automatically executing the flow of the computations for each conformation. The coupling of using Nimrod/G with program and script to extract the results, allows to fill the scoring function along the parameters sweeping and generates at the end an excel document with the final results.

4 Applications of Hybrid method on biosystems



4.1 Project 1: Engineered PLP-Dependant Aldolase



4.1.1 Motivations

This new described hybrid method was applied first to support experimentalists in their research to improve the stereo specificity of a retro-aldolase engineered enzyme. The project is lead by Professor Dr. Donald Hilvert from ETH and his graduate student Florian P. Seebeck and is about the conversion of a Pyridoxal 5'-phosphate (PLP)-dependent racemase into an aldolase **[6]**. Aldol reactions [137] are widely used in synthetic organic chemistry due to their proficiency in forming or breaking carbon-carbon bond under stereo selective control [138]. One aim of protein engineering is to generate enzymes with new or significantly improved activity. The challenge in the process of developing such enzyme is to be able to catalyze the desired chemical reaction. An aldol reaction involves a nucleophilic addition of an enolate donor to a carbonyl acceptor to form a beta-hydroxyl ketone. The aldol adduct counts 2 stereogenic centers on alpha and beta carbon position [139]. The control of those stereocenters are really important in pharmaceutical synthesis as identical molecular structures with different stereochemistry generate changes in chemical and biological properties. For example early research of Pfizer on heart disease has resulted by the marketing of a drug called Lipitor (atorvastatin) involving two aldol reactions in the synthesis process.

The retro-aldolase activity, as the name indicates, is used to cleave C-C bonds. Starting from the adduct aldol, the reaction mechanism is triggered by a strong base initiating the Cα-Cβ bond breaking by deprotonation of beta-hydroxyl group. Retro-aldol reaction exhibits substrate-based stereo-dependancy. The chirality on the carbon stereocenters, influences the stereochemistry of the outcome of the reaction. In particular a chiral center in the alpha position gives a good stereo-control of the enantio- and diastereo-selectivity of the reaction. Recent work have been reported on the development on new enzyme [140] or antibody [141] to catalyze in a useful scales retro-aldol reactions.

The investigations here consist in discovering which base initiates the retro-aldol reaction by studying all the possible conformations of the cofactor-bound-substrate (PLP-ß-hydroxy-α-amino acids) in the active site of the mutated enzyme (alr). The

experimentalists proposed Histidine(166) as a candidate to play the role of the base, but by our preliminary studies we discovered that an intramolecular reaction in the ligand could happen as the phosphate group of the cofactor can be accessed by ß-hydroxy moieties, depending on the conformations of the phenylserine (*figure 2*). The theoretical investigations will provide information regarding the optimal conformations of the ligand to fit in the binding site, more precisely the orientation of the phenyl-serine (substrate) into the pocket, and the associated electrostatic complementarity between the enzyme and the cofactor-bound-substrate, which provides mechanistic information about the reaction process. The computational results can also provide important insights for the experimentalists to make other mutations to improve the enantioselectivity of the enzyme.

4.1.2 Presentation of the bio-system of interest

The system of interest is the Pyridoxal 5'-Phosphate-dependent alanine racemase enzyme from *Geobacillus stearothermophilus*. The thermophilic characteristic of this bacterium is used to check sterilization cycles. The association of two dimers structurally builds the racemase enzyme. The reaction into the binding site of the racemase enzyme is initiating by a tyrosine residue 265. The experimentalists turn the racemase activity to an aldolase by a mutation in the binding site.

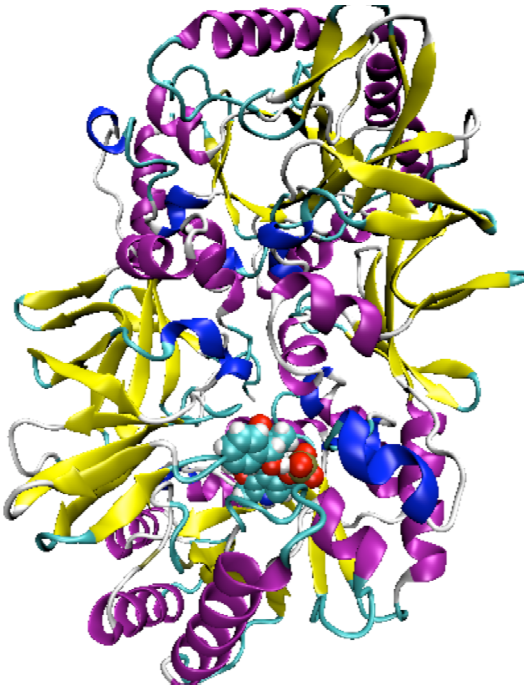


Figure 4-1 Alr Y265A enzyme reprensented with cartoon mode and PLP-phenylserine substrate modelized with van der Waals radii.

4.1.2.1 Active site description

The recombinant enzyme was produced by PCR and site-directed mutagenesis and purity was assessed by electrophoresis in the laboratory. The process results in a single amino acid substitution in the active site causing change of both substrate specificity and reaction profile, turning the **racemase** activity into an **aldolase** activity. The comparison between PLP-dependant alanine racemase with L-threonine aldolase from *Thermotoga maritima*, two evolutionary unrelated enzymes with different structures, substrates and reaction profile, revealed an analogy in the activation of their catalytic cycle. They both form an aldimine intermediate between PLP cofactor and their respective substrate: alanine in alanine racemase and ß-hydroxyl-alpha-amino acid in L-threonine aldolase. The aldimine is subsequently

converted to a quinoid intermediates by a proton abstraction on Cα in the case of alanine racemase and by general base-promoted cleavage of C_α - C_β bond in L-threonine aldolase. The retroaldol mechanism involved in the enzymatic reaction is described in Figure 4-2. Figure 4-3 illustrates the mutation carried on the binding site as well as the PLP bound phenylserine substrate. Figure 4-4 shows the ligand inside the active site.

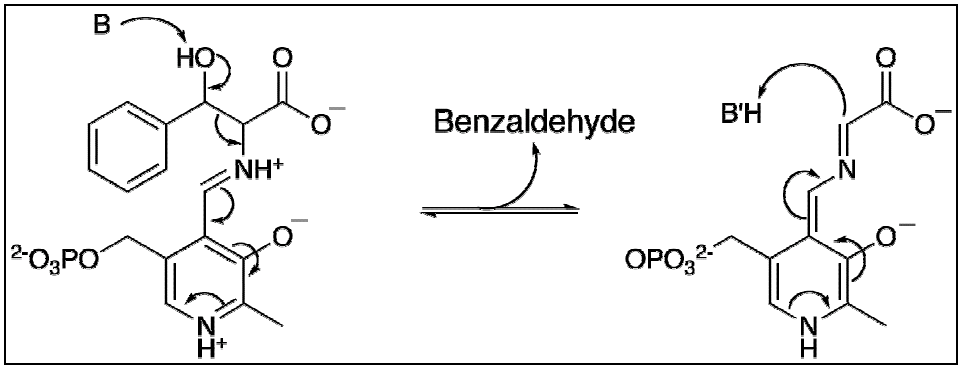


Figure 4-2 Base-promoted cleavage of the C_α-C_β bond on the cofactor and the substrate (phenylserine)

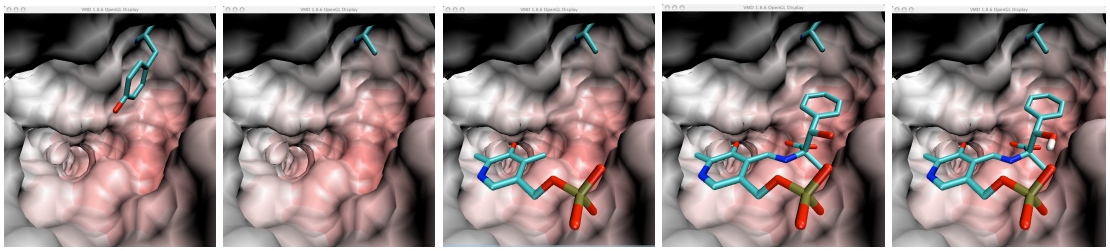


Figure 4-3 Active site of the enzyme represented as molecular surface.

From left to right, tyrosine 265 originally in the binding site of alanine racemase, Alanine 265 replacing tyrosine residue and increasing the space in the active site, PLP cofactor, phenylserine part added to the cofactor and entire ligand with the proton of interest.

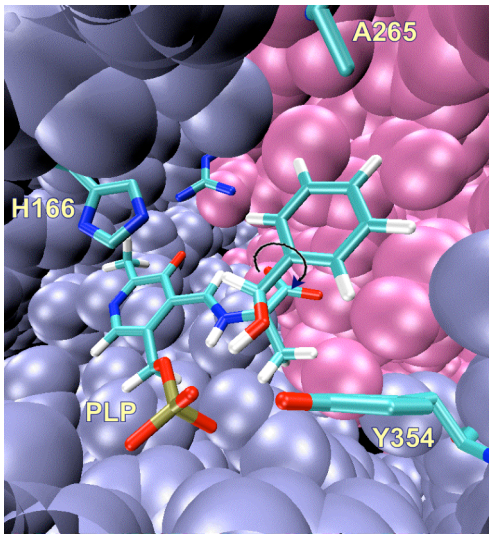


Figure 4-4 The active site of the enzyme (Subunit A in blue, Subunit B in mauve) with PLP bound to ß-hydroxy-α-amino acids

The experimentalist synthesized 4 diasteroisomers of the PLP cofactor bound to β-hydroxyl phenylserine. Only two conformers were showing activity in the binding site from a circular dichroism spectroscopy analysis.

4.1.2.2 Active substrate in the binding site

There exists two enantiomeric pairs of the ligand, (2R,3S) with (2S,3R) and (2R,3R) with (2S,3S). From the laboratory experiments it was reported that only 2 of the 4 isomers are showing activity in the binding site of the enzyme. The preferred conformation is the 2R- (Figure 4-6). This could be explained by looking at the conformation of the 2S- ligand sitting in the pocket. The model of the ligands shows steric incompatibility in the cavity for enantiomers 2S-, bumping occur with residue lysine 39 as shown on Figure 4-5. The steric obstruction is due to the reorientation of the phenyl ring of the ß-hydroxy-α-amino acid in the other side of the plane of PLP cofactor. 2S- ligand remains instable in the pocket, which explain the absence of activity in the active site of the enzyme.

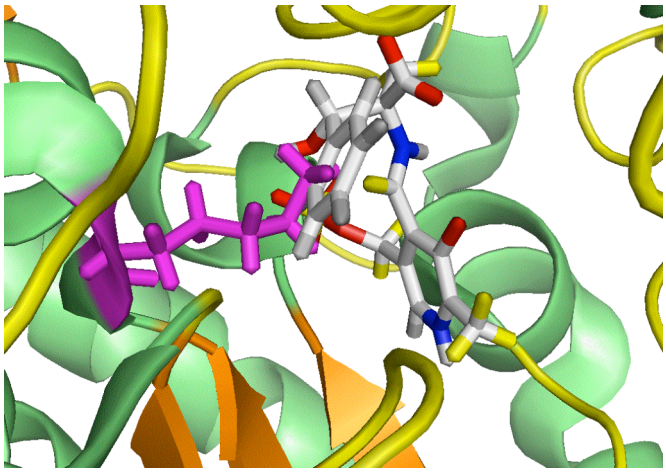


Figure 4-5 Isomer (2S,3S)-α-methyl-β-phenylserine in the binding site.

The 2S- isomer cannot fit into the binding site due to steric clash with the lysine 39 side chain represented in magenta.

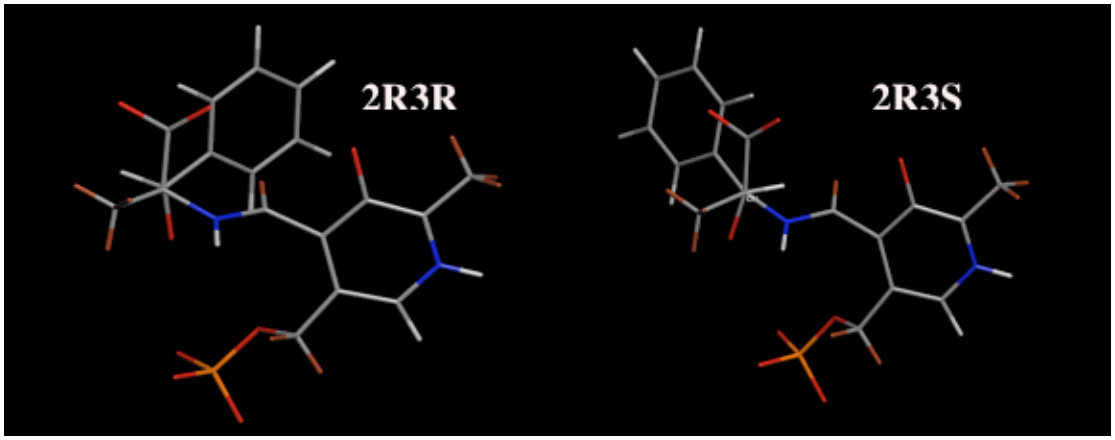


Figure 4-6 (2R,3R) and (2R,3S)-phenylserine substrates active in the enzyme.

The ligand requires a particular conformation to be able to react in the active site. The electronic rearrangement due to the proton abstraction implies the atoms between the Cα of ß-hydroxy-α-amino acids to the ring of the PLP cofactor align in the same plan. The cofactor is well anchored inside the active site of the enzyme by hydrogen bonding network. Those interactions stabilizing the cofactor rigidify its position in a way that let motions only around 2 particular torsion angles. The first torsion angle called kappa is located between the Cα - Cß. Bond of the serine and the second one between atoms Cß and C1 of the phenyl ring. The first torsion allowed the reorientation of the hydroxyl group carrying the proton of interest for the activation of the enzyme. The second torsion is important in the sense that it governs the investigation of the steric fitting of the phenyl ring in the pocket now increased in size by the mutation. Figure 4-7 below shows the two degrees of freedom (DOF) that will be applied on the ligand along the quest of the best conformation of the phenyl-serine bound to the cofactor scoring the highest relative binding energy.

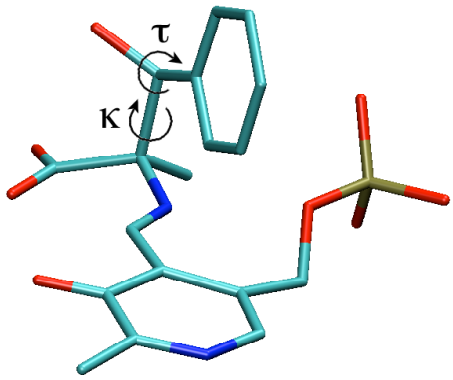


Figure 4-7 Kappa and Tau, the two torsion angles defined to investigate conformational accessibility of the two isomers of the ligand

4.1.3 Results

4.1.3.1 Protonation

4.1.3.1.1 Protein

The protonation of the protein is important in molecular recognition, especially around the binding site. The reaction occurring in this region depends strongly on the protonation state of hot key residues. The protonation state leads the interaction as donor-acceptor, draws the hydrogen bonding. In the enzyme of interest the protonation of the histidine 166 is really important factor as it was likely the initiator of the reaction. Histidine can adopt two tautomeric forms depending on the orientation of the imidazole ring [142]. This orientation is quite challenging to determine precisely by electron density with x-ray technique. In the structure retrieved from the PDB, the histidine amino acid appears in a protonated form in the delta position more often than on the epsilon site but it was showed that the hydrogen on the nitrogen epsilon is the most common form encountered in protein.

Using the tool pdb2pqr to add hydrogen atoms on the protein, the histidine side chain was flipped and the imidazole ring gets protonated on nitrogen Delta. The visualization of the protein with all the hydrogens in presence of the ligand tells us about the correct orientation of the ring and protonation state that favors the reaction the most. A hydrogen bond occurs between histidine 166 and arginine. So the binding or reaction rate depend a lot on the protonation state of the residue directly in contact with the ligand.

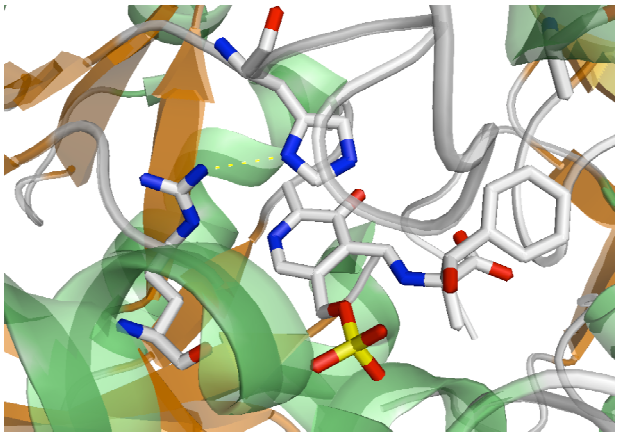


Figure 4-8 Hydrogen bond between HIS166 and ARG219 from the active site of the enzyme

4.1.3.1.2 Ligand

The hydrogen atoms on the ligand are added with the program babel and a visualization of the results needs to be carried out. First the ligand itself needs to be checked to determine if the amount of charges is respected and secondary it has to be visualized inside the pocket. From this view the key interactions stabilizing the complex should be checked and corrected if necessary. In our case the nitrogen atom of the ring of the cofactor was protonated but looking at the interaction with the enzyme, Arg^{219} is forming a hydrogen bond with the cofactor.

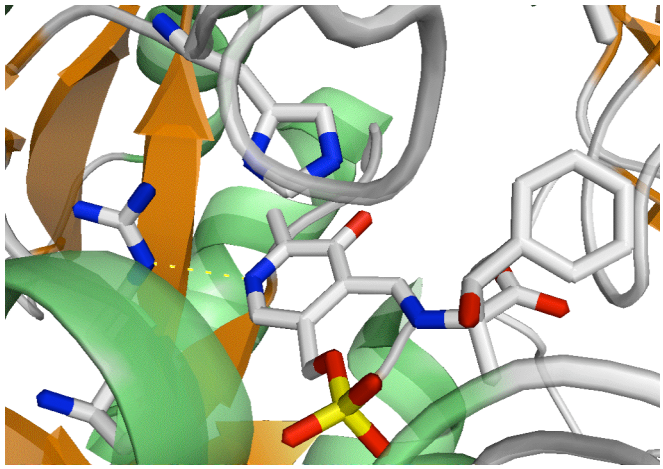


Figure 4-9 Hydrogen bond interaction between PLP-cofactor and Arg[219]

The positioning of the proton is arbitrary and not really optimized. As we have seen, the hydrogen atoms play an important role, so an optimization is performed at a B3LYP/DZV(2d,p) level of theory in a dielectric 4 to mimic the interior of the protein. The dielectric is applied by using the Polarization Continuum Model implemented in GAMESS.

4.1.3.2 Conformation accessibility of 2R3R

The torsion along tau and kappa are done on a full rotation of 360 degrees. For every 30 degrees of rotation on kappa, a full rotation along tau is done. The investigation consists in a matrix of 12 by 12 so 144 conformations. The torsion variations are done using the conformational tool analysis program, developed by Chris Kondric as mentioned in the hybrid method section. The program generates the new conformation of the ligand and checks if no clashes happen between the atoms of the ligand itself. In addition the program has a module to replace the ligand inside the protein and scan the distances of each atoms of the ligand with the atoms of the binding site to detect the ligand conformations that can not fit inside the binding site and so can not contribute to the activation of the enzyme.

From Figure 4-12, a set of good conformations of the substrate is highlighted between the angles of rotation corresponding to kappa 11 to kappa 3. The conformation showing intermolecular clashes can be taken in account due to the flexibility of the receptor occurring in biological environment. An example of intermolecular clashes is illustrated on Figure 4-10. A periodicity in the results depending on the torsion along tau appears due to the rotation of the phenyl ring. The distances between the hydroxyl group and the oxygen atom of the phosphate group decrease from kappa 2 to 6. The set of conformations in this range shows compatible distances to predict a potential proton abstraction from the phosphate group that will play the role of the base. The reasonable structures are concentrated in the range of short distance between the epsilon nitrogen of histidine 166 and the hydroxyl group.

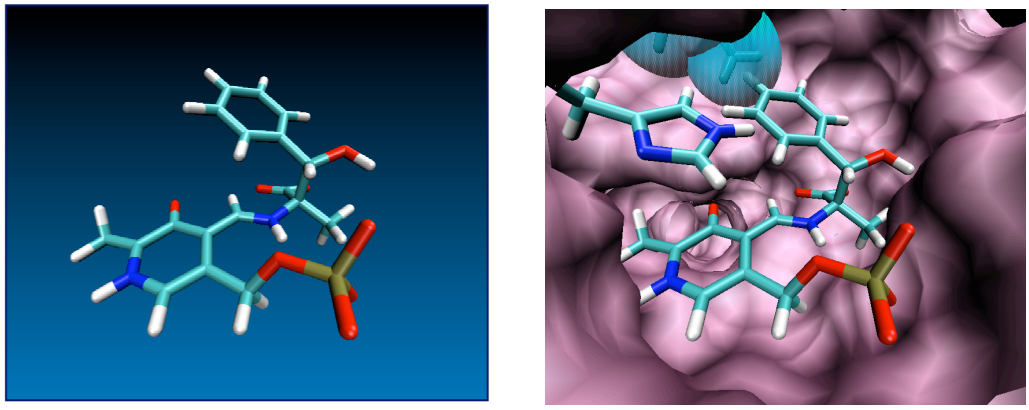
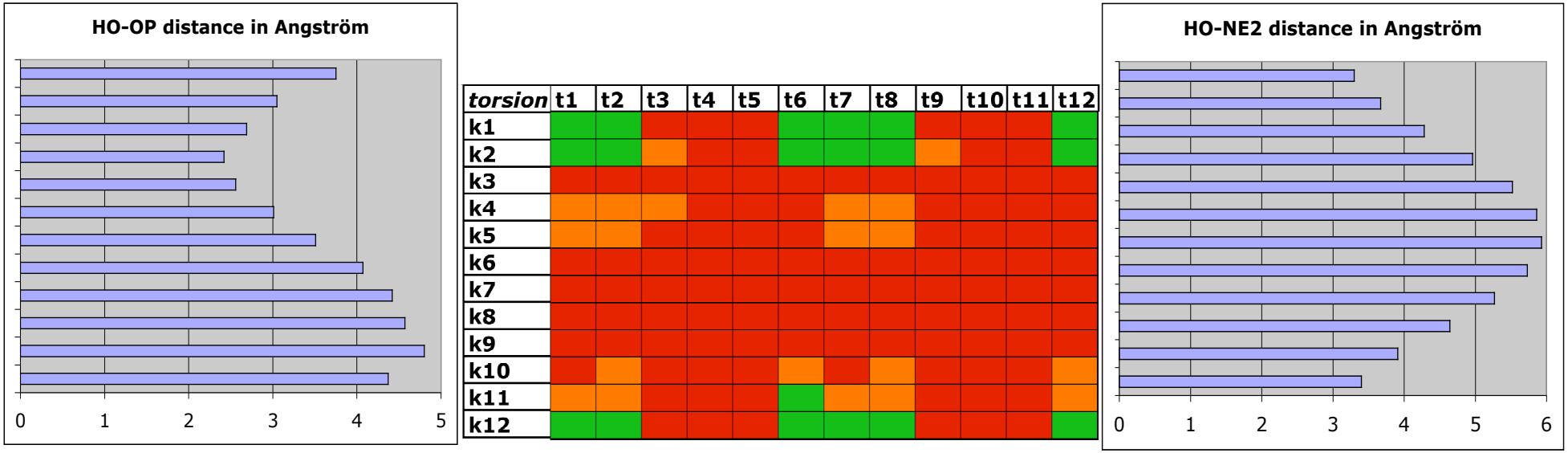


Figure 4-10 Ligand 3D representation on the left side and replacement of the ligand into the binding site on the right side.

This picture shows an intermolecular clashes happening in the pocket when structural changes are applied on the ligand.

Figure 4-11 PLP bound (2R,3R)ß-hydroxy-α-amino acid conformation accessibility

The central table shows the results from the conformational investigation along the two torsion angles kappa and tau defined. The red color indicates the conformations of the ligand replaced in the active site representing intra-molecular and intermolecular steric clashes, orange represents intermolecular bumping between atoms of the protein and atoms from the substrate, the green coloring indicates the absence of bad contacts. The graph on the left plots the distances between the oxygen atoms of the hydroxyl group and the Oxygen atom of the phosphate group in the ligand. The graph on the right plots the distance between the oxygen atom of the hydroyl group of the ligand and the epsilon nitrogen atom of the histidine residue from the active site of the enzyme.

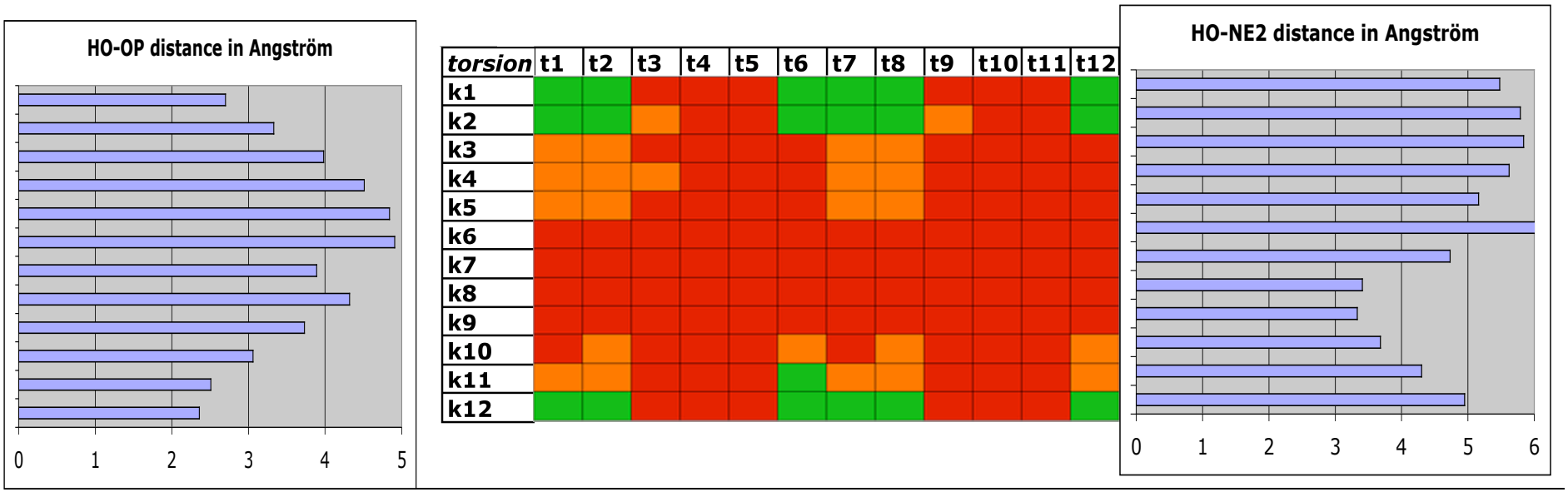


4.1.3.3 Conformation accessibility of 2R3S

The same treatment that was applied on (2R,3R) was also carried out on (2R,3S). The matrix of the docking results shows a range of good conformations between kappa 11 to kappa 3. The short distances between the phosphate group and the hydroxyl are situated in the same range as the feasible structures. The (2R,3S) ligand is more susceptible to interactions with the phosphate group than with histidine of the binding site. The distance between the oxygen atom of the hydroxyl group and the oxygen atom of the phosphate group varies between 2.3 and 3.3 Angström that represent reasonable distance for proton abstraction. The results are reported on Figure 4-12.

Figure 4-12 PLP bound (2R,3S)ß-hydroxy-α-amino acid conformation accessibility

The central table shows the results from the conformational investigation along the two torsion angles kappa and tau defined. The red color indicates conformations of the ligand replaced in the active site representing intramolecular and intermolecular steric clashes, orange represents intermolecular bumping between atoms of the protein and atoms from the substrate, the green coloring indicates the absence of bad contacts. The graph on the left plots the distances between the oxygen atoms of the hydroxyl group and the Oxygen atom of the phosphate group in the ligand. The graph on the right plots the distance between the oxygen atom of the hydroyl group of the ligand and the epsilon nitrogen atom of the histidine residue from the active site of the enzyme.



4.1.3.4 Internal energy

The stability of the new conformations of substrate generated with the geometry tool are estimated with the quantum chemistry package GAMESS. Ab initio computations are performed at B3LYP/DZV(2d,p) with constraints applied to the optimization. Only the hydrogen atoms are allowed to rearrange structurally. The Final energy corresponding to the minimum energy structure of the molecule, relates the internal energy of the molecule. The comparison of the different internal energy gives a good insight on the stability of one conformer to the other one. In Figure 4-13, the most stable conformers of 2R,3S is k0t0 with a set of conformations around like k0t30, k0t150, k30t0 and k30t30. Figure 4-14 reports the internal energy for the different configurations of 2R,3R. From the graph it clearly appears that the set of kappa 30 structures are the most stable conformations.

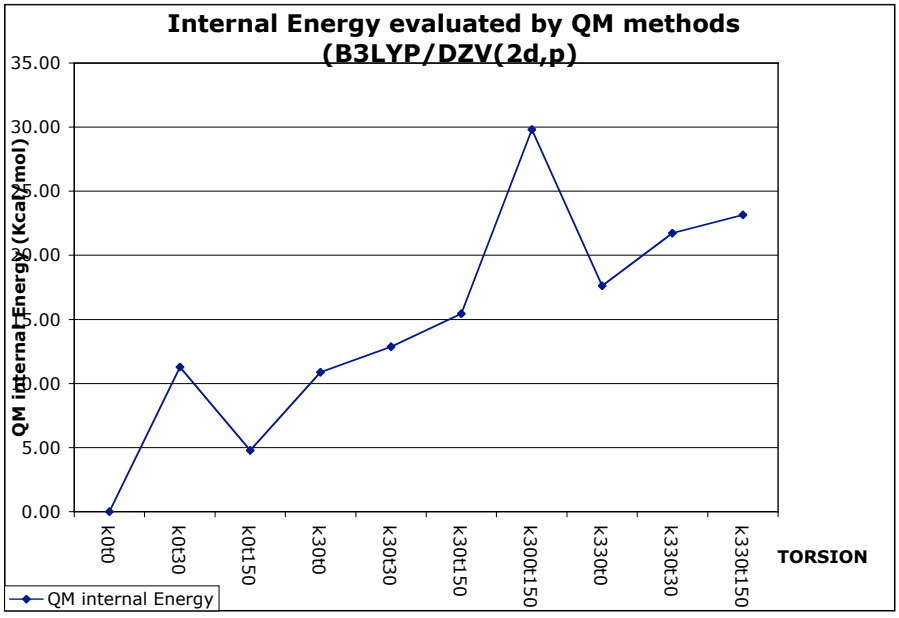


Figure 4-13 Internal energy of (2R,3S) ligand conformations, estimated with ab initio program GAMESS at B3LYP/DZV(2d,p) level of theory.

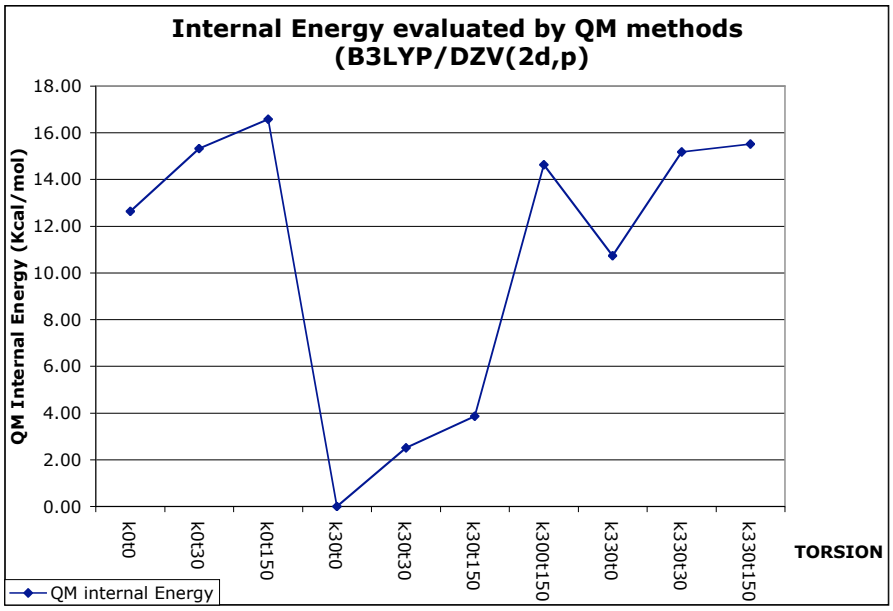


Figure 4-14 Internal energy of (2R,3R) ligand conformations, estimated with ab initio program GAMESS at B3LYP/DZV(2d,p) level of theory.

4.1.3.5 Binding energy

After the geometry analysis of each new enzyme/substrate complex, the structures with good steric complementarity are selected in order to estimate the binding energy. APBS package is used to calculate the electrostatic energy difference between the complex and the protein and ligand alone. The results are reported in Figure 4-15 for the conformer 2R,3S. The lowest binding energies values are concentrated for kappa value of 30, with k30t150, k30t30, k30t0, and for kappa 300 with k300t150. The estimation of binding energy for 2R,3R shows lowest energies around kappa 30 with k30t150, k30t30, k30t0 (Figure 4-16). For both conformers the same set of structures appears to have the lowest binding energy.

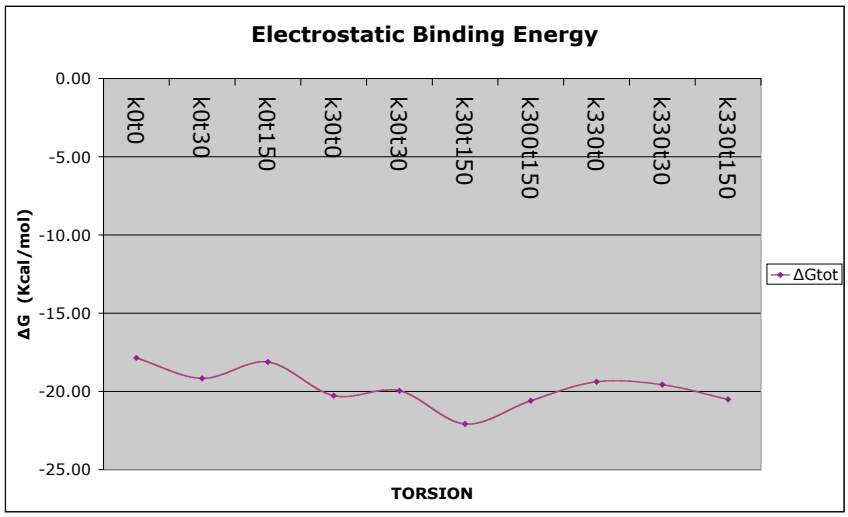


Figure 4-15 Binding energy in Kcal/mol between Alr Y265A and substrate conformations with (2R,3S)-phenylserine.

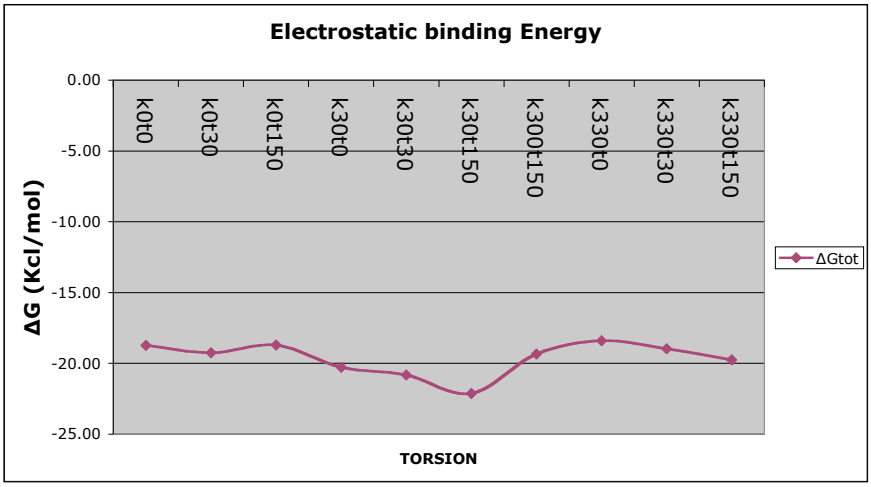


Figure 4-16 Binding energy in Kcal/mol between Alr Y265A and substrate conformations with (2R,3R)-phenylserine.

4.1.3.6 Scoring the potential candidates

The scoring function is applied on each conformation of the ligand to sort out the best binders between the different orientations of the phenylserine moiety bound to the PLP cofactor. As described in the hybrid method section, the scoring function combines binding energy of each complex with the internal energy of each substrate.

The total binding energy of (2R,3S) conformations still show lowest energy for k0t0, as reported on Figure 4-17. The full scoring function applied on (2R,3R) revealed still the same set of structure around kappa 30 like the binding energy and internal energy (Figure 4-18). The lowest conformations for (2R,3R) is k30t0.

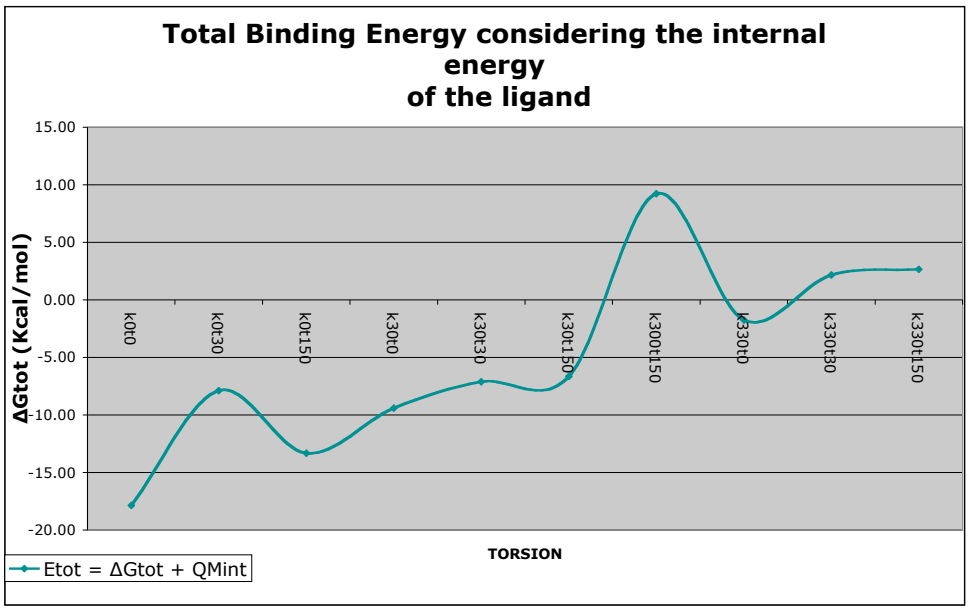


Figure 4-17 Total binding energy between (2R,3S)-phenylserine conformations bound to cofactor and Alr Y265A enzyme.

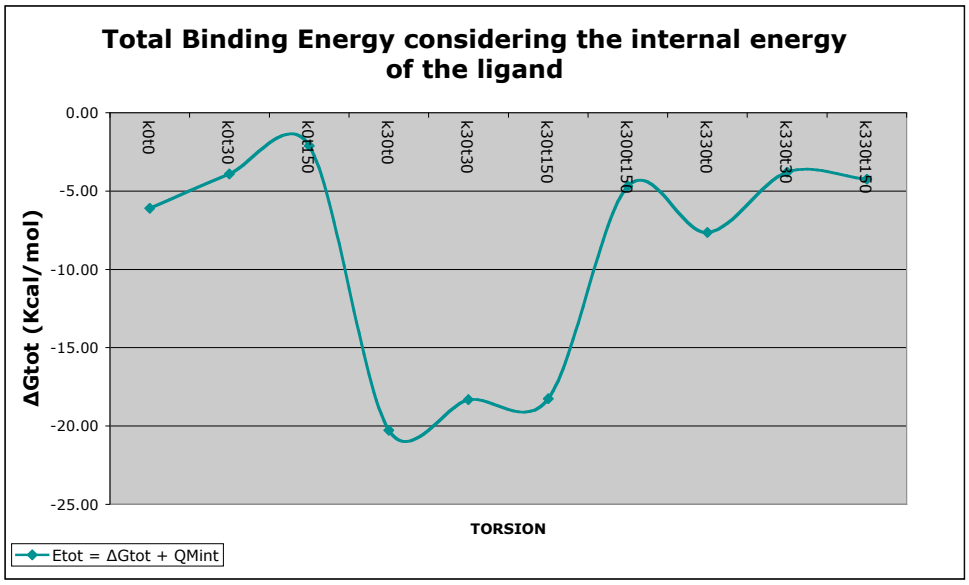


Figure 4-18 Total binding energy between (2R,3R)-phenylserine conformations bound to cofactor and Alr Y265A enzyme.

4.1.3.7 Orbital analysis

To understand the reaction occurring in the binding site, *ab initio* molecular orbital calculations were performed at second-order Møller-Plesset perturbation theory with double zeta valence basis set incremented by polarization functions. The reaction of interest involves a proton abstraction on the hydroxyl group attach to the carbon-α and the oxygen atom of the phosphate group. The reaction can occur if the sigma O-H bond overlap with the p orbital of the oxygen from the phosphate group. Both structures selected after the geometry and energetic analysis, have the hydroxyl bond pointed toward the oxygen attached to the phosphate atom. The orbital analysis confirms the existence of an overlap between OH group and the oxygen acting as a base (Figure 4-19 and Figure 4-20).

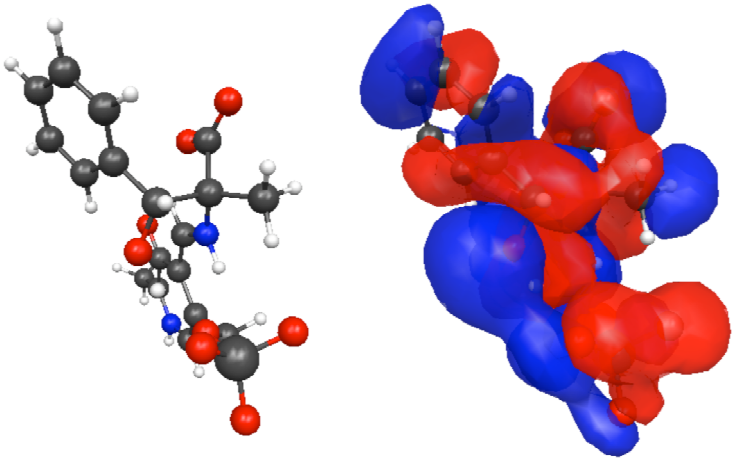


Figure 4-19 (2R,3R)-phenylserine in conformation k30t0 on the left and on the right Molecular Orbital map onto.

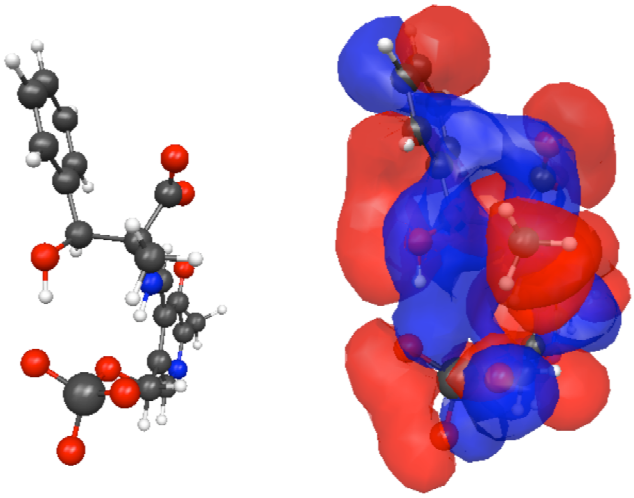


Figure 4-20 (2R,3S)-phenylserine in conformation k0t0 on the left and on the right Molecular Orbital map onto.

The alignment of both structure k30t0 and k0t0 highlight a close positioning of the phenyl ring in the pocket. This suggests a confined pocket with not much room for the ring to rotate or translate. The hydroxyl groups of both diastereoisomers are pointed to the oxygen of the phosphate in a just opposing way due to the chiral β-center. The distance between the two oxygen atoms, 2.7 and 3.05 for (2R,3S) and (2R,3R) respectively, is in agreement for hydrogen bond. The hydrogen atom is pointed toward the lone pair of the oxygen as shown on Figure 4-21.

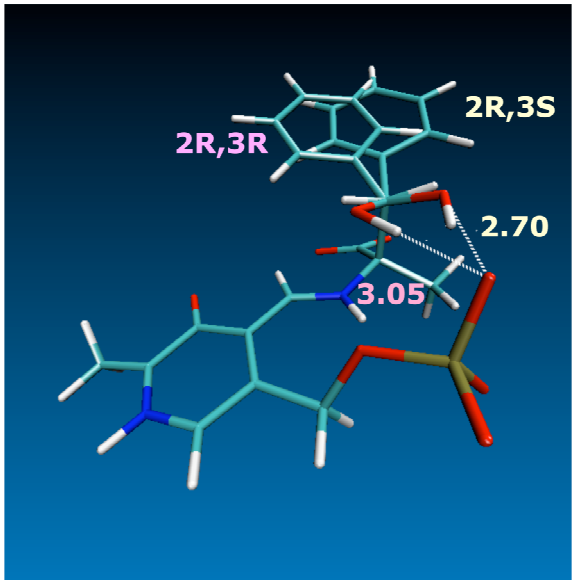


Figure 4-21 Alignment of (2R,3S) and (2R,3R) structure sort out after the investigation

4.1.4 conclusion

The investigations based on the different stereoisomers of α-methyl-β-phenylserine substrate bound to PLP cofactor show that only diastereoisomer 2R- can fit in the active site of the enzyme. 2S- α center can not fit without any clashes with residues of the binding site. The stereoselectivity for the β center is absent as either the 3R- or 3S- isomers can fit in the pocket with close binding energy values. The space given by the substitution of the tyrosine to alanine offers room to fit the phenyl ring of the substrate. The phenyl ring can not adopt a lot of position in the pocket and both diastereoisomers (2R,3R) and (2R,3S) adopt similar position, with the phenyl ring perpendicular to the PLP plan with a 30° of difference by a rotation along cα -cβ bond.

The initial assumption on the histidine 166 to play the role of the base to abstract the proton of the hydroxyl group was eliminated during the investigation. The histidine adopted a wrong tautomeric form. The imidazole ring flipped and the protonation state was on the delta position. Looking at the hydrogen networking it appears that the histidine engages a hydrogen bond with arginine 219 side chain, as shown on figure 20.

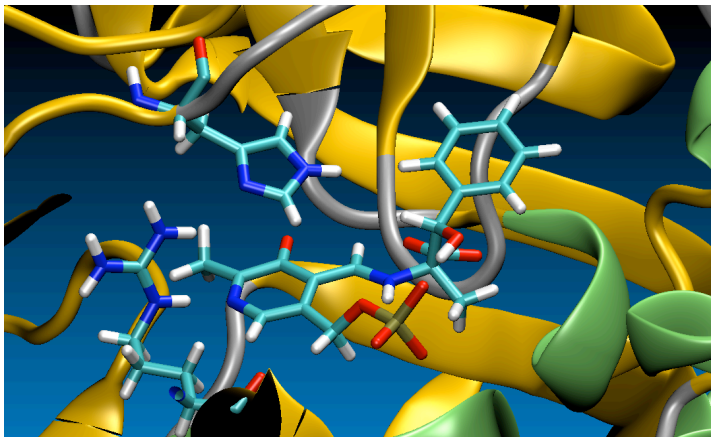


Figure 4-22 Substrate in the active site of the enzyme. Histidine 166 engaging hydrogen bond with side chain of arginine 219.

The second assumption regarding the phosphate group was proved to be the right mechanism. The oxygen atom attached to the phosphate play the role of the base to abstract the proton of the hydroxyl group. The hydroxyl group of both diastereoisomers is pointing toward the lone pair of the oxygen atom of the phosphate group. The distance between the two oxygen atoms is favorable for hydrogen bonding in both cases as shown in Figure 4-23.

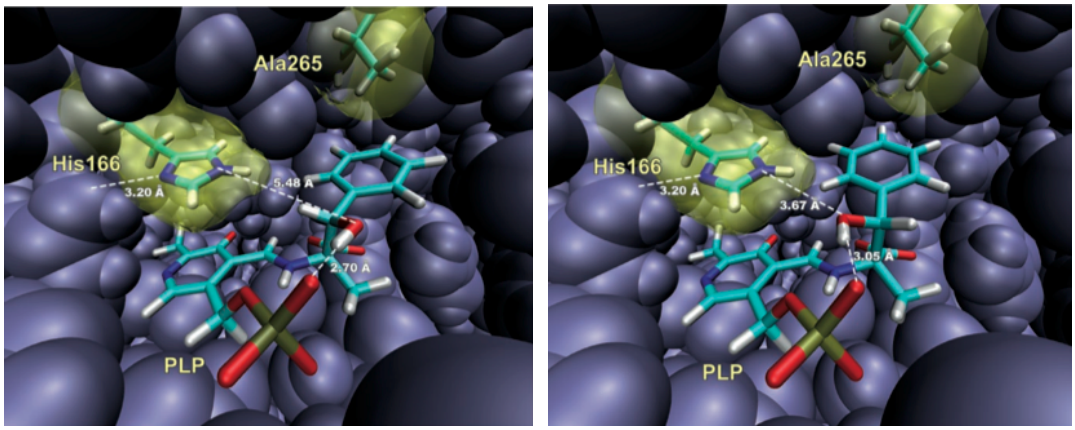
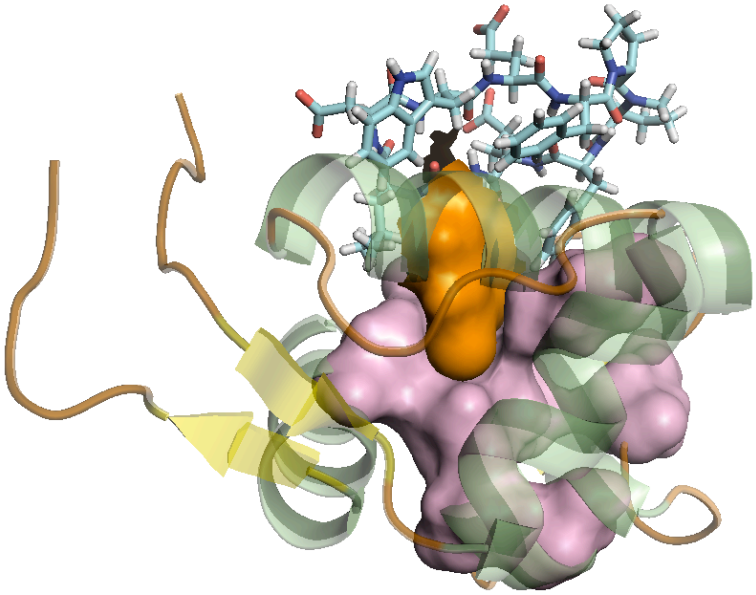
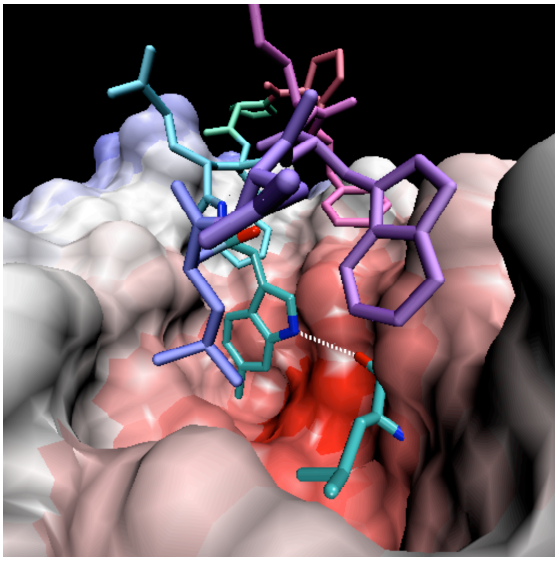


Figure 4-23 The two best conformations of the substrate, (2R,3R) and (2R,3S) different only by 30º from the carbon β.

The hybrid method was successfully applied on this system and brings some good insight to the experimentalists.



4.2 Project 2: β-hairpin Protein Epitope Mimetic Inhibitors of p53-HDM2 Protein-Protein Interaction



4.2.1 Motivations

Today advanced research in oncology focuses on a Protein-Protein-Interaction (PPI) drug target composed by the p53 tumor suppressor and its negative regulator the human double-minute2 HDM2 receptor. P53 is a transcription factor involved in DNA reparation, cell cycle arrest and activation of the apoptotic program. It is qualified as 'guardian of the genome'[1] inducing cell death to stop the progression of a tumor. The goal of a tumor is to proliferate into the cells and not be stopped by a "security agent", so the malign cells find a way to thwart the functions of p53 by mutating p53 gene or by over expressing the gene coding for HDM2 receptor. HDM2 inactivates and initiates the degradation of p53 offering the eternity to the cancerous cells. The strategy of the tumors is to avoid any accumulation of p53 active into the cell.

One therapeutic answer consists in reintroducing the functionality of p53 into the cell, in the presence of p53 mutant. Gene therapy can be applied to reintroduce the functional gene coding for p53 into cancerous cells. Another therapeutic target is to prevent the interaction between HDM2 and p53. The later strategy, PPI inhibition is the one investigate here. PPI strategy has a great interest in therapeutic research because of the structural diversity of the receptors and the specificity of the ligand to act on it. In drug discovery the investigations are focus on small molecule (ligand, peptide) and not protein because they are more favorable in terms of Biopharmaceutical properties (ADME: Absorption, Distribution, Metabolism, Excretion) to be able to pass the biological membrane and to not be recognized as an antigen inducing an immune response. In HDM2 receptor, hot-spot residues drive the binding interaction with the trans-activation domain of p53. In laboratory, the development of small inhibitors to mimic the part of the p53 domain involved in the binding process is an active research axis. Helped by structure-based design and combinatorial chemistry the experimentalists work together with theoreticians to build the most potent and efficient drug to bind HDM2 receptor.

4.2.2 Review of research done on antagonist of p53/HDM2

Several antagonist peptides have been discovered that bind HDM2 receptor (Figure 4-24) and it is still an active research area of drug development [143, 144]. All the peptides initiate the conformational change on the receptor upon binding and yield a good efficiency of the drug. Robinson and coworkers designed another kind of peptide using a cyclic beta-hairpin scaffold in 2006 [145]. In this peptidomimetic, hydrogen atom in position 6 on indole ring of the tryptophan 23 is substituted to a Chlorine atom (Figure 4-25). This substitution was proved to increase the binding between MDM2 receptor and 8-mer p53 peptide analogue [146] released by Novartis in 2006.

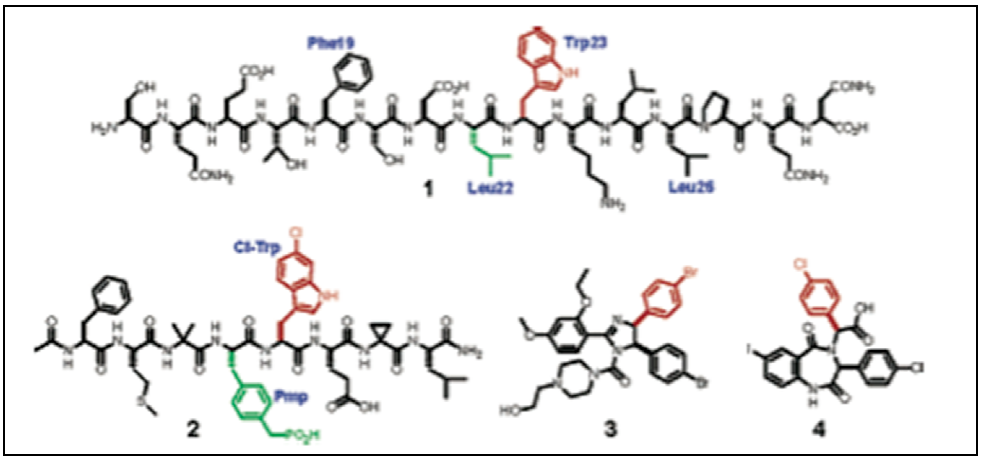


Figure 4-24 Structure of MDM2 inhibitors [146].

1- linear15-mer peptide (1ycr). 2- 8-mer peptide analogue (2gv2) from Novartis. 3- imidazoline inhibitor (1rv1). 4- benzodiazepine (1t4e)

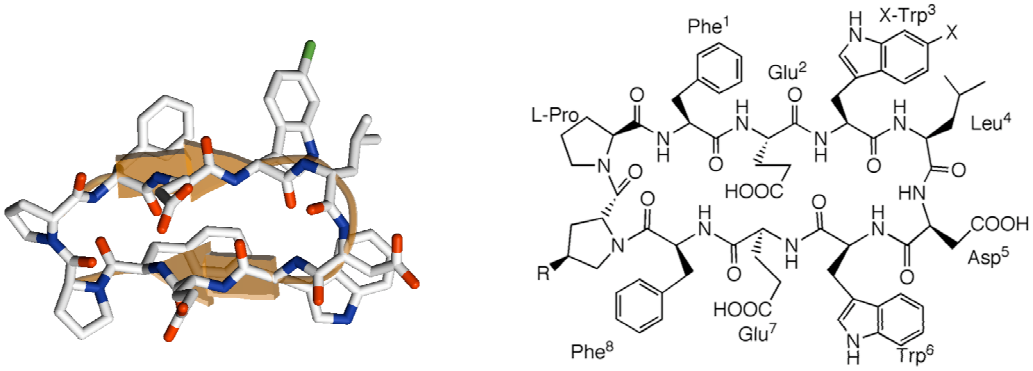


Figure 4-25 Beta-hairpin conformation of decapeptide synthetized by Robinson group.

Theoretical investigations on the interactions between MDM2 oncoprotein and truncated peptide from the N-terminal α helical part of p53 tumor suppressor has been carried out. The group of Professor Peter Kollman applied the MM-PBSA/GBSA [147] computational method with an alanine scan on each amino acid of the p53 peptide and highlighted the key residues involved in the binding. They especially they showed the importance of W23 in the binding energy. They also discovered that there was room left between two phenyl rings of the binding site and W23 from the peptidomimetic. They filled this empty space by replacing hydrogen atom on the indole ring to a methyl group in position 6 and 7. The substitutions show an improvement in the binding energy.

In Professor Robinson's lab, Dr. Anja Graesli applied the alanine scan technique experimentally on the beta-hairpin form of the transactivation domain of p-53. The experimental results gave the same as the theoretical conclusion from the MM-PBSA/GBSA method regarding the key residues.

4.2.3 Structural characteristics of p53-HDM2 interaction

4.2.3.1 Inside the binding site

The x-ray structure of the complex MDM2/p53 reveals an interface consisting in the recognition of three hydrophobic residues (Phe19, Trp23, Leu26) within the 15 amino acids of the p53 transactivation domain. The binding site of the receptor HDM2 appears as a deep hydrophobic cleft when it enters in contact with p53. The binding pocket is composed of 14 hydrophobic amino acids, which establishes several vdW contacts with p53. The buried residues Phe19, Trp23 and Leu26 of p53 mainly cover them. In addition two intermolecular hydrogen bonds are established, one at the entrance of the cleft between Gln72 of HDM2 and Phe19 of p53 and the second one between HDM2 Leu54 and p53 Trp23. Those two important interactions are conserved in 15-mer linear peptide whereas the cyclic form of the peptide only conserved the Trp23 (Trp3) interaction.

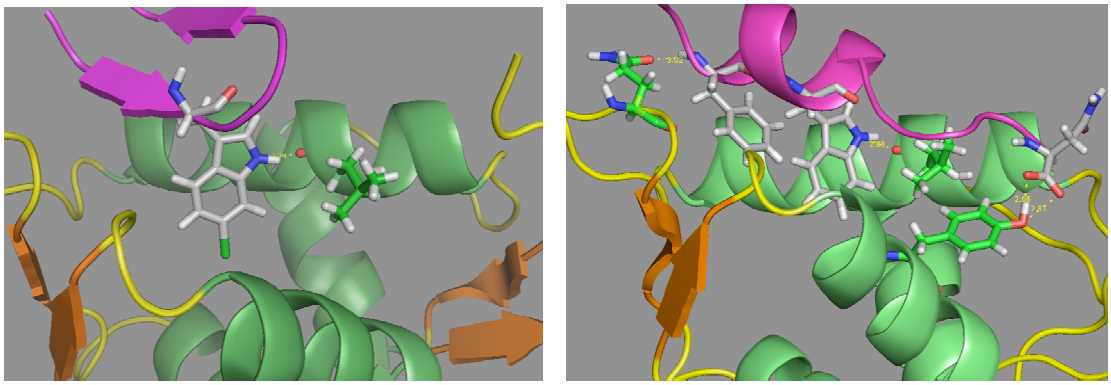


Figure 4-26 Hydrogen bond occurring in the binding site of HDM2 receptor.

The picture on the left shows the cyclic peptide with the hydrogen bond between the tryptophan23 and the leucine 54 of the binding site. The picture on the right highlights the two hydrogen bonds between phenylalanine 19, tryptophan 23 from the linear peptide with glutamine 72, leucine 54 respectively.

4.2.3.2 The inhibitor side

4.2.3.2.1 Linear peptide

15 residues from the transactivation domain of p53 constitute the linear form of the peptide. It was introduced in the protein Databank in 1996 under the identifier 1ycr [148]. The Figure 4-27b shows the large surface of the peptide embedded into to the binding site. The crystal structure of the linear peptide contains the wild type tryptophan. This amino acid is pointed toward to aromatic residues, F86 and F91, of the binding pocket. The orientation of the ring is more in alignment with F91 as shown on Figure 4-27c.

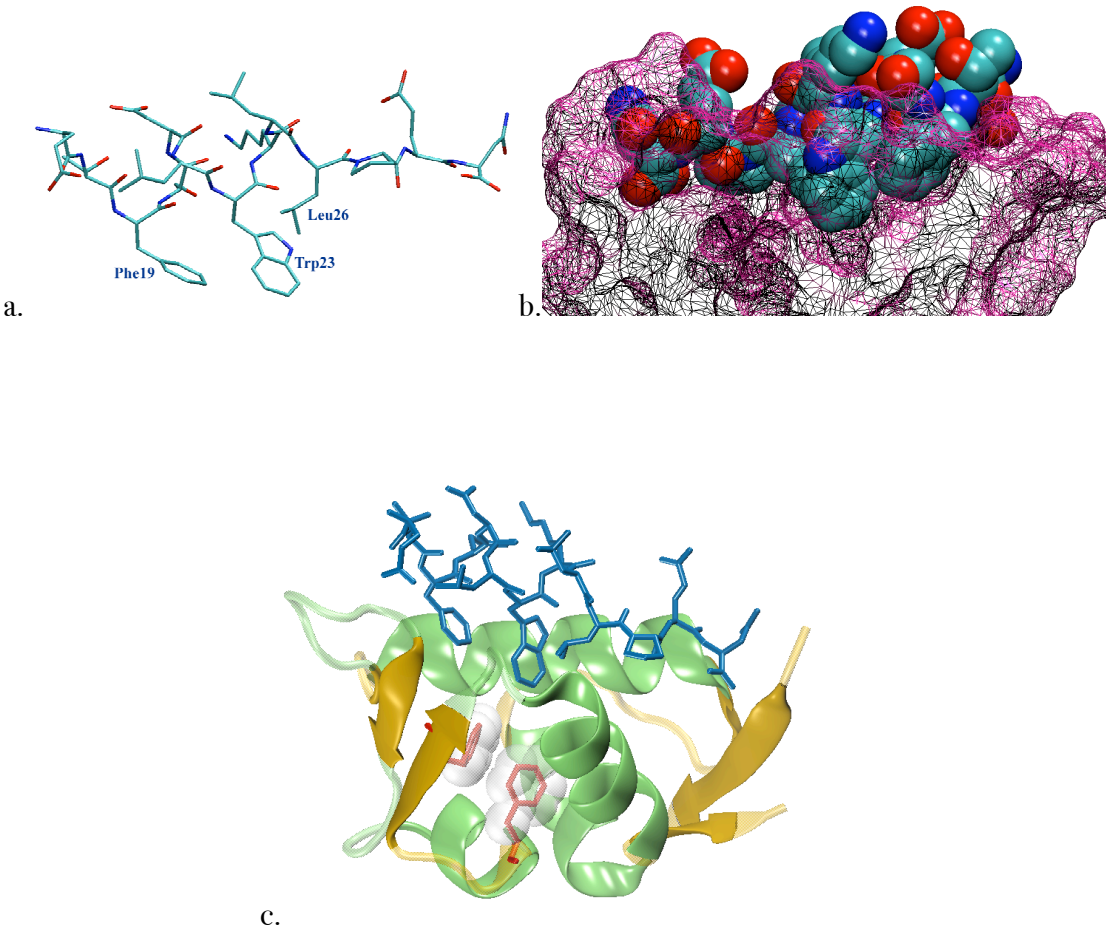


Figure 4-27 Linear peptide from 1YCR PDB structure

a. Linear peptide composed by 15 amino acids from the transactivation domain of p53 on the left.

b. Surface of the receptor represented as wireframe and van der Waals representation of the peptide.

c. Cartoon representation of the receptor with F86 and F91 shown as licorice supplemented with transparent van der Waals surface. The linear peptide is represented in blue with the tryptophan 23 pointed toward the aromatic ring of F86. The two phenylalanine adopt a T-shape orientation.

4.2.3.2.2 Cyclic β-hairpin peptide

A cyclic beta-hairpin peptidomimetic was reported in PDB recently under the identifying 2axi [145]. The cyclic peptide contains W23 (6Cl-W3) forming the interface between p53 and HDM2 receptor. The peptide was optimized compared to the first β-hairpin proposed in 2004 [149]. With addition to aromatic residues, the cyclic peptide becomes more stable. A Chlorine atom was added in position 6 of the indole ring of the Trp23 to improve the binding affinity. This substitution effect was shown on the linear peptide [150].

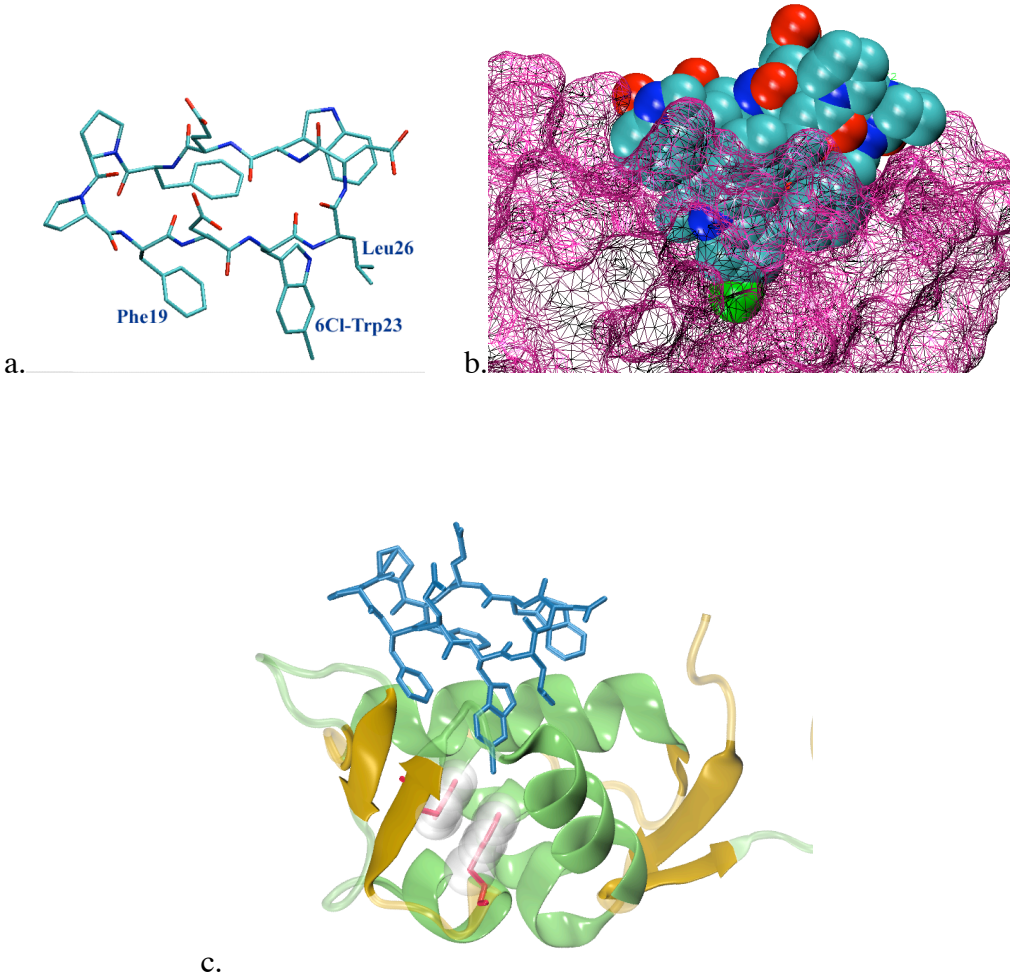


Figure 4-28 Cyclic β-hairpin peptide

a. Licorice representation of the cyclic peptide.

b. Surface of the receptor is represented as wireframe and peptide described as van der Waals surface.

c. Cartoon representation of the receptor with F86 and F91 shown as licorice supplemented with transparent van der Waals surface. The linear peptide is represented in blue with the 6-Chloro-tryptophan 23 pointed to an edge of the aromatic ring of F86. The two phenylalanine residues adopt a parallel stack orientation.

4.2.4 Guide lines of the project

Computational studies were undertaken using quantum mechanical methods, with the goal of characterizing the interactions between HDM2 residues forming the Trp^{23} binding pocket and the 6-chloro- and 6-methyl- as well as un-substituted Trp side chains (referred to below as X-Trp, where X = Cl, Me, or H) in linear (**13**-**15**) and cyclic (**6**-**8**) peptide ligands (

Table 4-1). The key X-Trp side chains are situated close to aliphatic (Ile^{57}, Ile^{61}, Val^{93}, Ile^{99}, and Ile^{103}) and aromatic (Phe^{86} and Phe^{91}) side chains in HDM2 (Figure 4-29).

CYCLIC PEPTIDOMIMETICS (2AXI)	X
6	H
7	Cl
8	Me

LINEAR PEPTIDE (1YCR)	X
13	H
14	Cl
15	Me

Table 4-1 label for 2AXI and 1YCR peptides and the three substituents

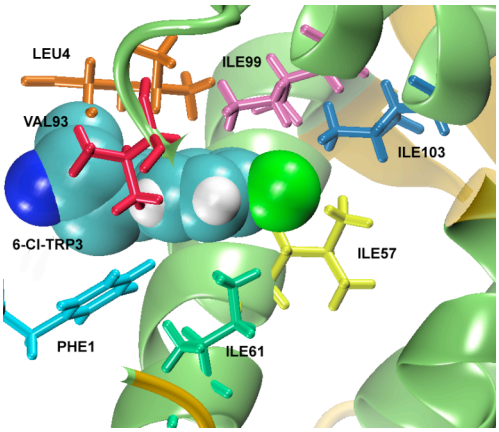


Figure 4-29 X-Trp residue in the protein surrounding

Of primary interest are; 1) understand the effect of the substitution of the H on the 6-C of the indole ring by a chlorine atom and then a methyl group 2) the orientation and strength of the hydrogen bond between the indole NH of X-Trp (peptide) and Leu54 (HDM2), and how this is influenced by the substituent X; and 3) the interactions between the X-Trp side chains in the ligands and the aromatic phenyl side chains of Phe86 and Phe91 in HDM2.

4.2.5 Substituent effect

The study concerns the variation of the charges inside the rings to better understand why the interaction of the 6-chloro-tryptophan in the binding site of the HDM2 oncogene receptor shows more inhibitor activity than the regular Trp or 6Me-Trp. For each form of the peptide, Trp^{23} differs only in the X substitution of the C6 of its indole ring (X=H, X=CH3, X=Cl). The computational study aims to provide more details on the structure and properties of each ligand as well as the interaction of each ligand within the pocket. The ligands considered here, involve a core tryptophan structural unit (Figure 4-30).

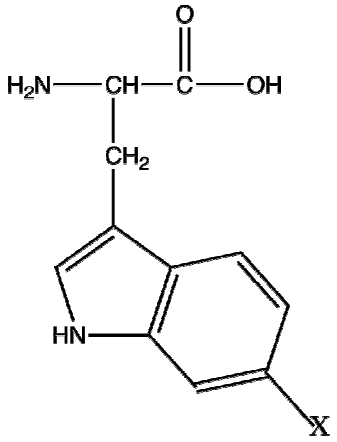


Figure 4-30 Tryptophan structural unit

In order to understand the effect of the substituent on the electronic structure, a sequence of quantum mechanical structure and property analyses were performed, including the following:

1- benzene, chlorobenzene, toluene

2- indole, 6-chloroindole, 6-methylindole

3- tryptophan, 6-chlorotryptophan, 6-methyltryptophan.

The quantum chemistry package, GAMESS, was used to study electrostatic properties of the different structures. The computations were done using the three Becke's parameters (B3) with the Lee, Yang and Parr (LYP) correlation functionals (B3LYP). The basis set used is double zeta valence increased in flexibility by adding polarization functions: DZV(2d,p). In order to estimate the effect of the substituents on the structure, the optimized structure of the molecule substituted by hydrogen was taken as a reference and then the substitution where done and a geometry optimization was performed only on the substituent moiety, by freezing the coordinates of the other atoms.

The atomic charges are computed using the fitting methods ChElPG [115] with Møller-Plesset perturbation theory and double zeta valence basis set with polarization functions (2d,p). ChElPG schema reproduces the electrostatic potential of the molecule by centering point charges on atom. The electrostatic potential surface is visualized with the program Molekel [151]. The molecular electrostatic potential is mapped onto the surface of the total electron density and for consistent comparison the same scale of electrostatic potential energy values is selected for each set of compounds studied.

4.2.5.1 Benzene

4.2.5.1.1 Molecular Electrostatic Map

Addition of halogen to benzene directs electron density to ortho- and para- positions through resonance effect due to the lone pair donation. This effect is weak compare to its electron withdrawing effect through an inductive effect on sigma frame of the arene ring due to the electronegativity difference (2.5 for the carbon and 3.0 for the chlorine atom). Results of the electrostatic potential maps illustrate the ortho- and para- directing tendency as compared to the benzene core unit (Figure 4-31). Alternatively, methyl substituent group weakly activates the ring through resonance donating effect in ortho- and para- position. In general, the methyl-substituent is a stronger ortho- and para- director than the chlorine atom, due to its donating inductive effect through sigma frame. Halogen and Methyl substituent effect can be compare to the regular benzene with its conserved symmetric distribution of electrostatic potential.

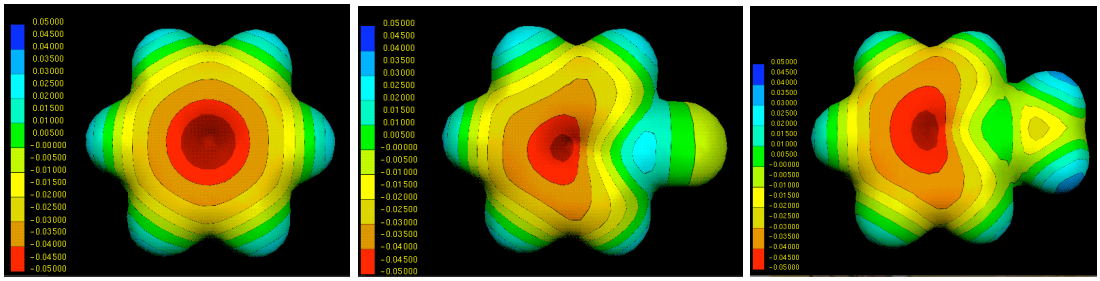


Figure 4-31 Molecular Electrostatic Potential Map of, from the left to the right, benzene, chlorobenzene and toluene

4.2.5.1.2 Electron Density Difference Map

Another way to evaluate the differences in electron density distribution with chlorine atom or methyl group substituted on benzene is to make a density difference between the substituted benzene. Figure 4-32 shows the results of such density difference analyses. In these plots, the ortho- directing tendency of the substituents is highlighted. The chlorine atom clearly shows less electron density in para- position compared to the methyl, due to an inductive electron withdrawing effect through the sigma frame.

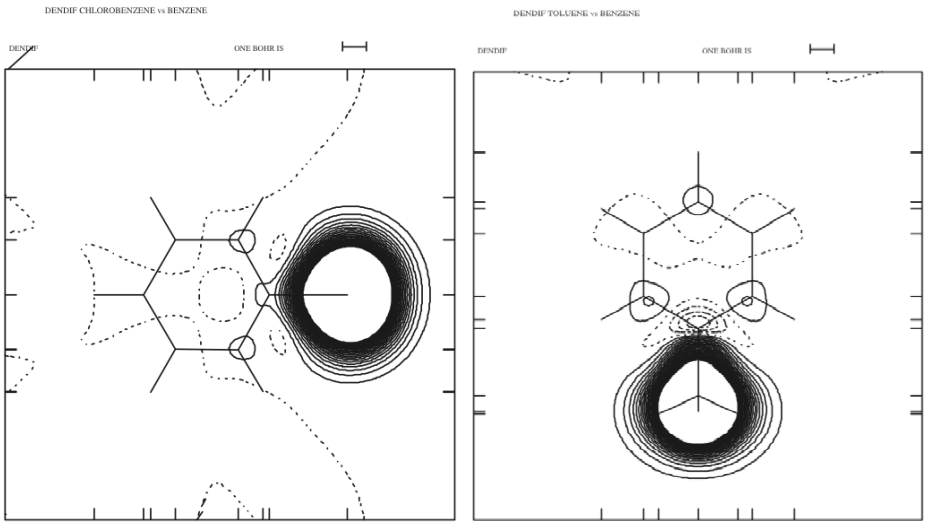
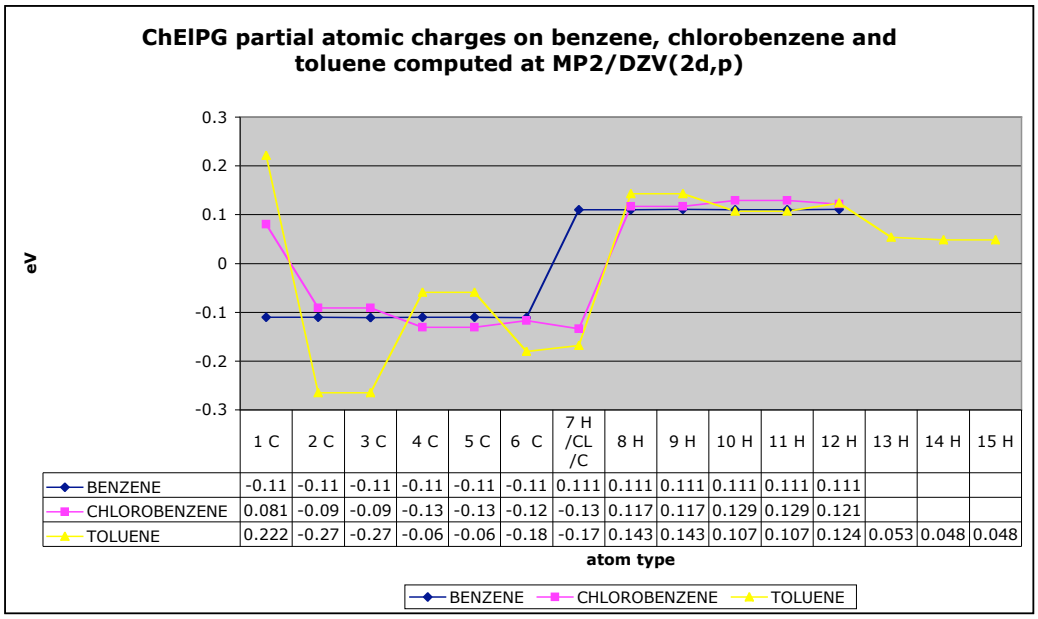


Figure 4-32 Electronic density difference map between chlorobenzene versus benzene at 0.7Å above the plan of the ring (left) and toluene versus benzene (right).

The full line represent the gain in electron density causing by the substituent and the dash lines describe the lost.

4.2.5.1.3 Partial atomic charges analysis

The substituent influences the partial atomic charges distribution. In the benzene molecule the distribution is symmetric whereas on chlorobenzene and methylbenzene, the charges are distributed according the effect of the substituent on the ring. As mentioned methyl and chlorine are ortho- and para-directors.



	1 C	2 C	3 C	4 C	5 C	6 C	7 H /CL /C	8 H	9 H	10 H	11 H	12 H	13 H	14 H	15 H
BENZENE	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11	0.111	0.111	0.111	0.111	0.111	0.111			
CHLOROBENZENE	0.081	-0.09	-0.09	-0.13	-0.13	-0.12	-0.13	0.117	0.117	0.129	0.129	0.121			
TOLUENE	0.222	-0.27	-0.27	-0.06	-0.06	-0.18	-0.17	0.143	0.143	0.107	0.107	0.124	0.053	0.048	0.048

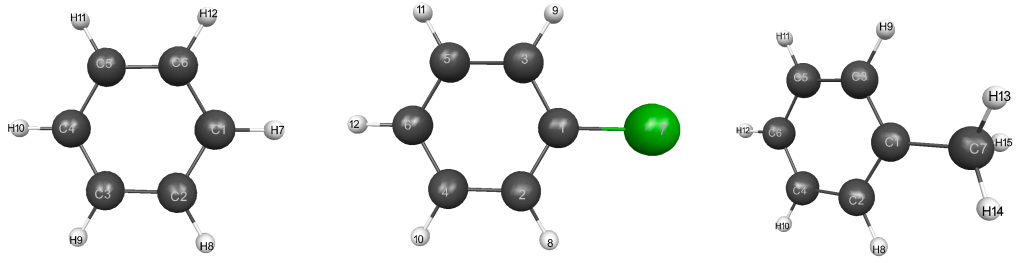


Figure 4-33 Plot of Chelpg charges computed at MP2/DZV(2d,p) level of theory for benzene, chlorobenzene and toluene.

The pictures at the bottom shows the atom numbering used to compare the charges of each molecule.

4.2.5.2 Indole

The indole takes part in the side chain of the tryptophan amino acid and we know from the study of the binding site that it interacts with the carbonyl group of the C terminal of residue Leu[54] by forming a hydrogen-bond. The analysis will show the effects of the substituent on the phenyl ring combined with the nitrogen atom of the pyrolle ring. The same study applied on the benzene was driven onto indole, 6-chloro-indole and 6-methylindole molecules.

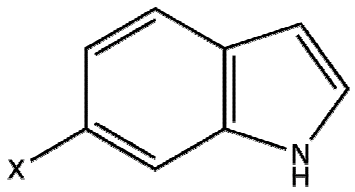


Figure 4-34 indole structural unit

4.2.5.2.1 Molecular Electrostatic Potential

In order to estimate the effect of the substituents on the indole ring, the optimized structure of the indole was computed at RHF//B3LYP/DZV(2d,p) level of theory. The optimized structure is the reference state and the different substituents were added on position 6 of the indole ring to replace the hydrogen atom. Geometry optimizations were performed at same level of theory and basis set with constraints allowing only the substituent piece to move. The resulting structure were use to map the electrostatic potential and compare the influence of the different substituents on the indole ring (Figure 4-35).

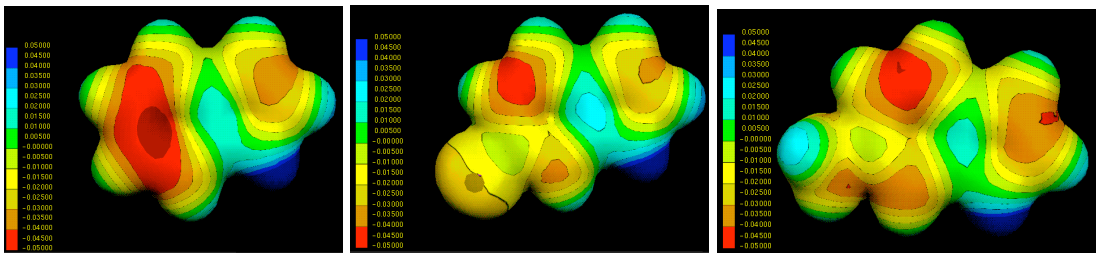


Figure 4-35: Molecular Electrostatic Potential map onto the electronic density of, from the left to the right, indole, 6-chloro-indole and 6-methyl-indole

Both the methyl and the chlorine, as we have seen in the study of benzene and its substituents, are ortho- and para- director groups stabilizing the phenyl ring by providing resonance. The methyl group has for effect to push the negative charge on the pyrolle ring and to give more nucleophilicity on the nitrogen atom, which reinforces the bond with its hydrogen atom. The chlorine atom has a withdrawing effect through sigma frame that decreases the distribution of negative charge on the pyrolle ring. This has for consequence to let the nitrogen sharing its lone pair with the ring and so more positive charges are spread on the nitrogen, with a more positive partial charge on the hydrogen. The hydrogen bond between the N of pyrolle ring and the Oxygen atom of L54 from the binding site becomes stronger.

The methyl group characteristically brings partial positive charge on the hydrogen atoms and more negative around the carbon atom. The chlorine atom has a small partial negative charge due to the sharing of its lone pair to the ring. Accurate study on the halogen show an electrostatic potential positive on the alignment of the carbon-chlorine bond whereas the side stay relatively negative [152]. This has the consequence of exposing to the hydrophobic pocket, a surface close to the neutrality on the chlorine atom with a slightly positive charges aligned on the C-Cl bond. The methyl substituent exposes more partial positive charge on the hydrogen to the phenyl ring of the phenylalanine 86.

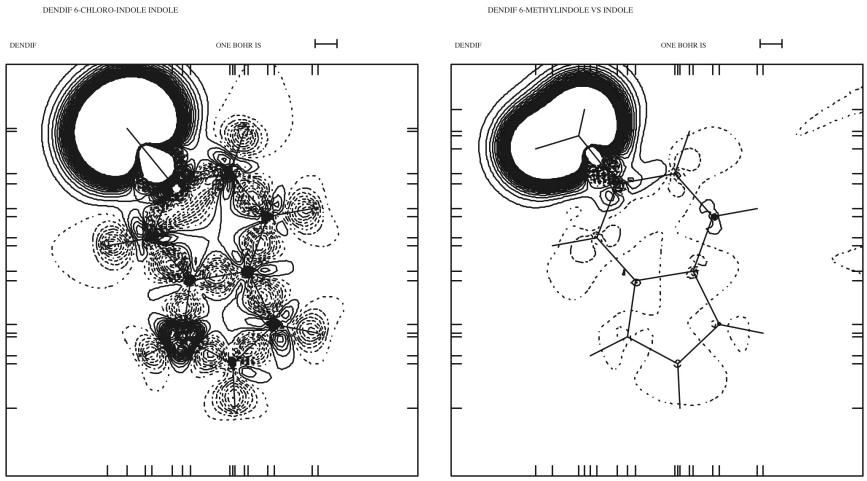


Figure 4-36 Electronic density difference map through the sigma frame (in the plan of the ring) between 6-chloroindole versus indole (left) and 6-methylindole versus indole (right)

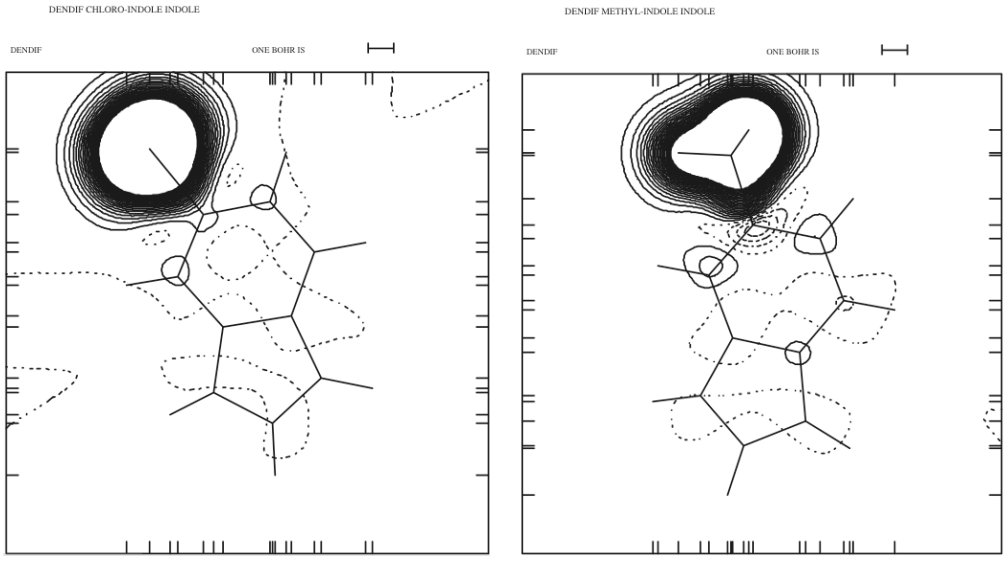


Figure 4-37 Electronic density difference map through the pi-frame (at 7Å above the plan of the ring) between 6-chloroindole versus indole (left) and 6-methylindole versus indole (right)

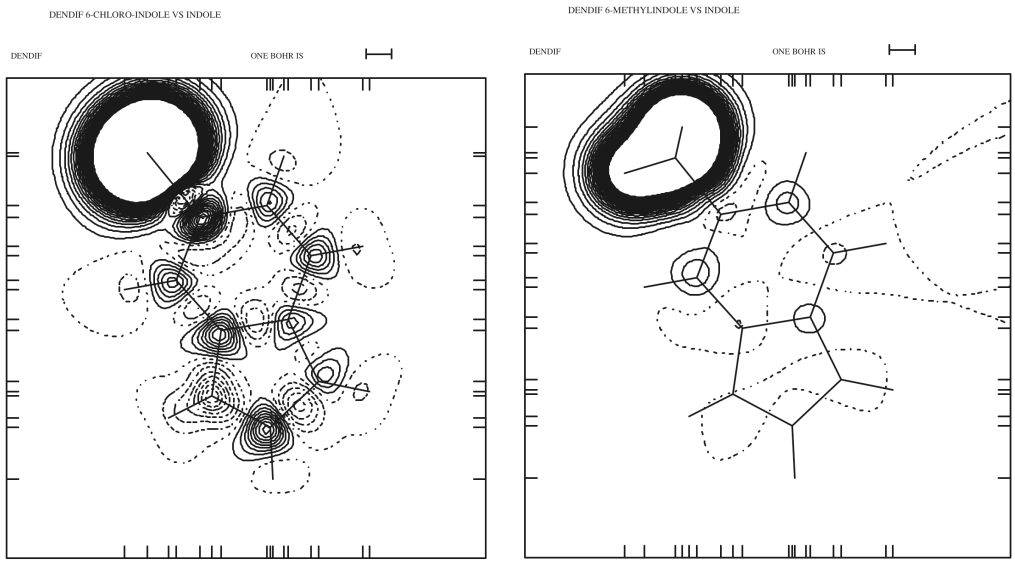


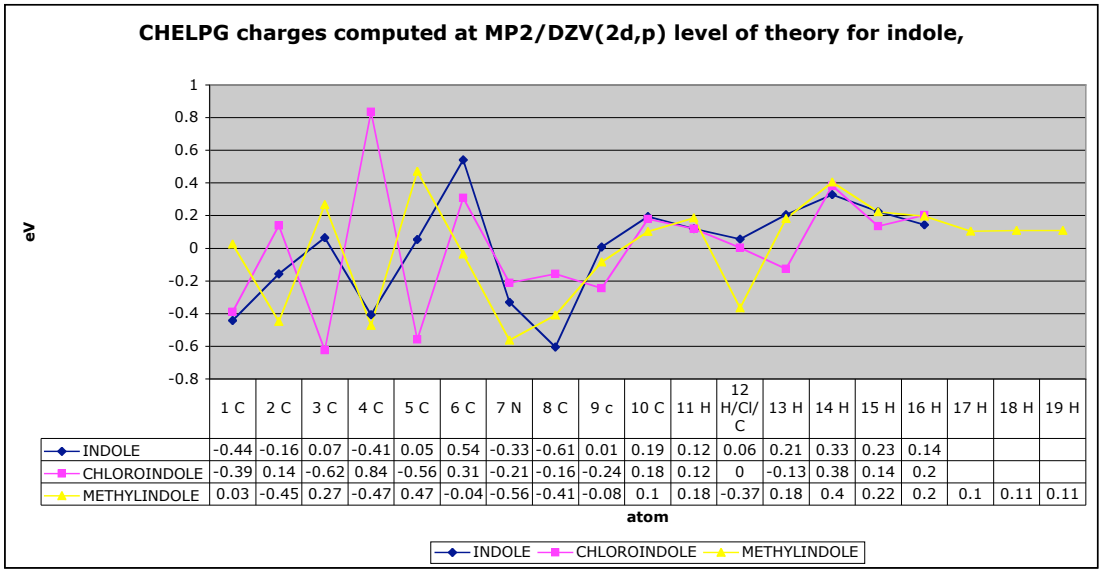
Figure 4-38 Electronic density difference map through the pi-frame (at 6Å above the plan of the ring) between 6-chloroindole versus indole (left) and 6-methylindole versus indole (right)

The halogen atom has a negative inductive effect through the sigma-frame as mentioned earlier; this can be seen on the density difference map when the map is drawn on the plane of the ring (Figure 4-36). The electronic density subtraction between indole and 6-methylindole cancels whereas the chlorine pulls the electron density on the ring except on the nitrogen atom on which the electron density is increased.

The density difference map drawn at 7.0Å above the plan of the ring (Figure 4-37) describes what happens in the π-system. The electron density is directing to ortho- and para- position for the phenyl for both substituents. By drawing the difference in electronic density at 6.0Å abode the plan of the ring (Figure 4-38) shows a net decrease in electronic density on the amine group (represented by dash lines).

4.2.5.2.2 Partial atomic charge analysis

The atomic charges are computed using the fitting methods CHELPG at MP2/DZV(2d,p) level of theory. The Figure 4-39 shows the difference on the charge of the nitrogen atoms related to each substituent. The nitrogen atom is less negative when chlorine atom is in position 6 of the ring compared to the hydrogen atom or methyl group that are increasing the negative charge.



	1 C	2 C	3 C	4 C	5 C	6 C	7 N	8 C	9 c	10 C	11 H	12 H/Cl/C	13 H	14 H	15 H	16 H	17 H	18 H	19 H
INDOLE	-0.44	-0.16	0.07	-0.41	0.05	0.54	-0.33	-0.61	0.01	0.19	0.12	0.06	0.21	0.33	0.23	0.14			
CHLOROINDOLE	-0.39	0.14	-0.62	0.84	-0.56	0.31	-0.21	-0.16	-0.24	0.18	0.12	0	-0.13	0.38	0.14	0.2			
METHYLINDOLE	0.03	-0.45	0.27	-0.47	0.47	-0.04	-0.56	-0.41	-0.08	0.1	0.18	-0.37	0.18	0.4	0.22	0.2	0.1	0.11	0.11

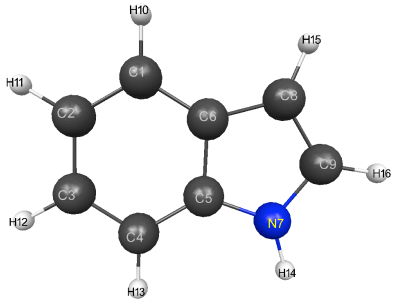


Figure 4-39 Plot of Chelpg charges computed at MP2/DZV(2d,p) level of theory for indole. 6-chloroindole and 6-methylindole versus indole

The study on benzene and indole structure provides good insight toward the understanding of how different types of substituent on the ring can influence the charge distribution. The results give information on how the molecule can interact in the binding site. The effect of the substituent can favor a hydrogen bond for example by pulling out the electron density on the donor group that will give more acidity to the nitrogen and favor the interaction of hydrogen with the acceptor group.

4.3 Investigation on the h-bond between Leu54 and Trp23

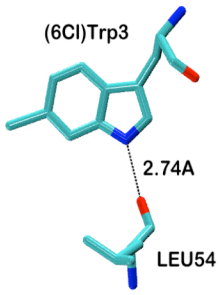


Figure 4-40: 6-chlorotryptophan from the cyclic peptide and Leucine54 from the binding site

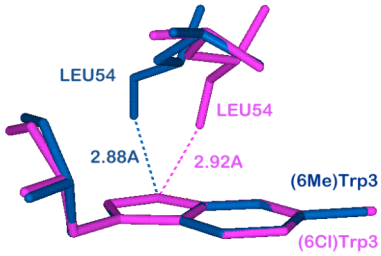
The distance between the amino group of the Trp^{23} indole ring and the carbonyl of Leu^{54} backbone is 2.74Å according to the crystal structure of the complex with the beta-hairpin peptide. As described before this h-bond is really important in the stabilization of the peptide during the binding process[145]. As we see in the previous chapter, changing the hydrogen atom in position 6 of the indole to a halogen or a methyl group, affect the distribution of partial atomic charge over the ring. Here the work focus on giving an estimation of the energy related to the h-bond under the different substitution.

4.3.1 How the substituent affect positioning and strength of the h-bond between Leu54 and Trp23

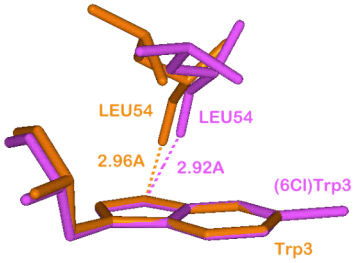
First, unconstrained geometry searches of the X-Trp-NH to Leu^{54} H-bond were performed on fragments composed of X-Trp plus Leu^{54}, starting with the interaction geometry seen in the 2AXI crystal structure and using density functional theory methods (M06-2X/DZV(2d,p)). This was done to explore the optimal structural orientation, charge distribution, and associated hydrogen bond strengths, in not only the crystal structure, but also in related energy minimized geometries with X=H, -Cl and -Me. The calculations were performed in the gas phase, and within a protein-like environment with a dielectric of 4. A dielectric of 4 is typically used to mimic the interior of a protein, although some estimates consider the dielectric can be higher. The resulting complexes are compared in Figure 4-41, which illustrate how the X-substituents influence the optimized hydrogen bond in terms of position, bond length, and therefore, bond strength. The hydrogen bond strengths were computed using both the crystal structure geometry, as well as the structures found by the unconstrained geometry optimization. The results are summarized in Table 4-2.

Figure 4-41 Structures for Leu[54] of HDM2 plus X-Trp of peptidomimetic, shown after optimization starting from the geometry in 2AXI, in the three 6-substitution forms:

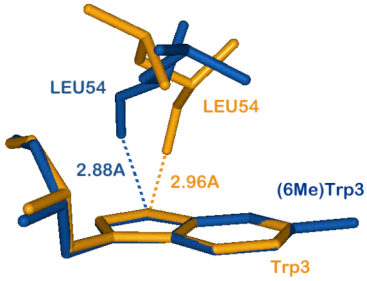
(a) overlay of X=Cl (pink) and X=CH3 (blue), (b) overlay of X=Cl (pink) and X=H (yellow), (c) overlay of X=Me (blue) and X=H (yellow). Also shown (d) is an overlay of Cl-Trp taken from the crystal structure (2AXI, grey) with that from the unconstrained geometry search (pink); also shown (e) are interactions of Cl-Trp (in 2AXI) with side chains of HDM2 (Ile57=yellow, Ile61=green, Ile99=pink; Val93=red, and Leu4=orange from the ligand).



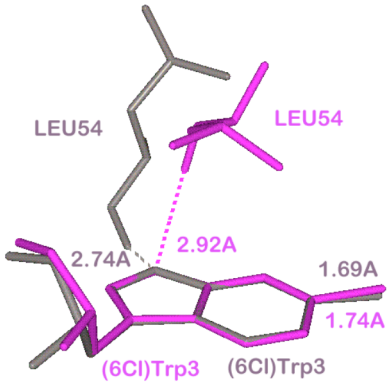
(a)



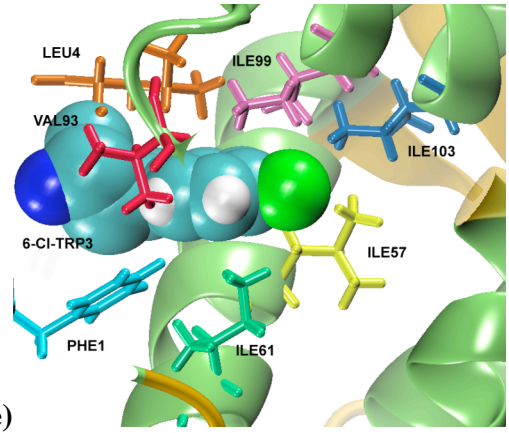
(b)



(c)



(d)



(e)

SYSTEM 2AXI + X-TRP	H-BOND ENERGY, DIPOLE MOMENT, & N---H----O DISTANCES[1,2]				
	Energy (Kcal/mol)		Dipole (D)	H-bond distances	
	ε=1	ε=4		N-O	N—H/H--O
X=H	11.4 (4.5)[3]	6.5 (2.1)[3]	5.0 (3.8)[3]	2.96	1.02 / 2.00
X=Cl	10.6 (5.2)[3]	5.8 (2.5)[3]	2.7 (5.0)[3]	2.92	1.01 / 2.06
X=Me	11.0 (4.6)[3]	6.4 (2.3)[3]	1.9 (4.9)[3]	2.88	1.01 / 2.06

[1]M06-2X/DZ(2d,p) fully optimized complex and fragments.

[2]N---H distance (N—H distance; H---O distance)

[3]Calculated using the crystal structure geometry (2AXI), with only the substitutions X=H, CH_3 and the resulting hydrogen bonds optimized with QM: H-bond length (N..O) = 2.74 Å.

Table 4-2 Calculated H-bond energies (kcal/mol) and distances between Leu54 of HDM2 and X-Trp in the cyclic peptidomimetics 6-8. Data are shown for the optimized geometries and (in brackets) for the crystal structure (2AXI) geometry.

4.3.1.1 Molecular Electrostatic Map

The molecular electrostatic map of substituted Trp[3] with Leu[54] from the binding site shows the influence of the substituent on the hydrogen bond between the two fragments (Figure 4-42). The Chlorine atom localizes more positive partial charge around the nitrogen atom compare to the methyl or the hydrogen substituent, which is in agreement with the observation made on the MEP map of the indole molecule. The hydrogen bond interaction between the carbonyl and the amino group has the effect of decreasing the partial negative charge on the oxygen atom in the fragment containing the chlorine substituent. This reflects a stronger hydrogen bond interaction initiated by the effect of the chlorine substituent.

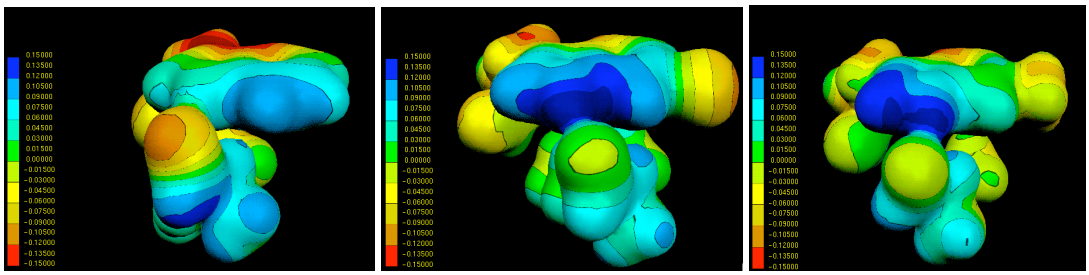


Figure 4-42 Molecular Electrostatic Map of the fragments, X-W23 and L54, for X=H, X-Cl and X=CH3 from left to the right

4.3.2 How the orientation of F86/F91 side chains influences on the binding

As mentioned above, the key ligand X-Trp side chains are in close proximity to aromatic residues Phe[86] and Phe[91] in the binding pocket on HDM2 (Figure 4-27c.). In the **7**-HDM2 crystal structure (2AXI), alternative side chain conformations were detected for both Phe[86] and Leu[57] by X-ray crystallography. In one orientation, the Phe[86]/Phe[91] side chains are in a slightly off-set parallel-stacked orientation with

respect to each other (dihedral 2.4°), and related perpendicular to the indole ring of the ligand, with the phenyl hydrogens pointing towards the chloro-substituent of the peptide ligand (Figure 4-43). In the alternative conformer, the Phe^{86} aromatic ring has rotated through ca. 34° to a tilted-stacked position with respect to Phe^{91}. In the complex of HDM2 with linear peptide **13a** (1YCR), the Phe^{86} is rotated further towards a T-stacked orientation, resulting in a dihedral angle of 55° between phenyl rings of Phe^{86} and Phe^{91}.

The Phe^{86} and Phe^{91} side chains are close enough to the X-Trp side chain of the peptide ligand to be influenced by the electronic nature of the substituent (X=Me or Cl). An unconstrained optimization, starting with the 2AXI structure, using a fragment comprising X-Trp with the two Phe^{86}/Phe^{91} residues provides an estimate of this interaction energy, which can be compared to the interaction energy, determined using the crystal structure conformation.

A stabilizing (electrostatic) effect on the complex is expected when X-Trp is X=Cl (**7**), due to an interaction between the rather diffuse chlorine atom and the partial positive charge of the Phe^{86}/Phe^{91} phenyl hydrogen atoms. Computations predict this interaction to be worth 4.0 kcal/mol for the optimized geometry structure in the gas phase, but <1 kcal/mol in a low protein-like dielectric (ε=4). In the crystal structure conformation, the interaction is computed as 2.3 kcal/mol in the gas phase, which is reduced to 1.4 kcal/mol in a protein-like environment with a dielectric ε=4. The calculated distances found for the chlorine to phenyl interaction in the optimized structures are similar to that found in the crystal structure Table 4-3.

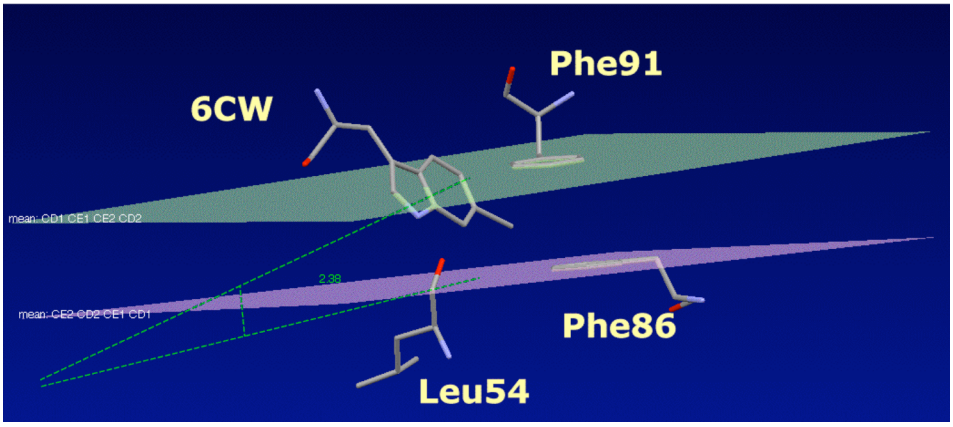


Figure 4-43 Dihedral formed by the intersection of the plan of the phenyl ring of Phe91 and the phenyl ring of Phe86.

Table 4-3 Calculated interaction energies (kcal/mol) between Phe^{86}/Phe^{91} of HDM2 and X-Trp in the cyclic peptidomimetics (6-8)**. Data are shown for the optimized geometries.**

TRP-X WHERE	E (KCAL/MOL)		SHORTEST X TO PHE HYDROGEN DISTANCES (Å)[1,2]	
	gas phase	ε=4	X---Phe^{86}-H	X---Phe^{91}-H
X=H	5.1	1.7	2.7, 2.8	2.2, 2.4
X=Cl	4.0	<1	3.3, 3.1	2.9, 2.8
X=Me[3]	2.6	<1	2.7, 3.0	2.3, 2.4

[1]X---Phe^{86} distance of 2 closest Phe H's (X—Phe91 distance of 2 closest Phe H's)

[2]Shortest distances found in crystal structure are 3.4 Å and 3.6-3.7 Å.

In the case of the methyl substitution (X=Me) there is the possibility for vdW interactions between hydrogens of the Me group and hydrogens of the Phe^{86} and Phe^{91} side chains. Computations predict an interaction energy of 2.6 kcal/mol in the gas phase, but <1 kcal/mol in a protein-like environment with a dielectric ε=4. For the case of X=H, computations predict a rather high interaction energy (5.1 kcal/mol) in the gas phase due to a considerable geometry change upon optimization in the absence of the protein environment, which reduces to only 1.7 kcal/mol in a protein-like dielectric (ε=4). Alternatively, when the conformation is held to that of the crystal structure, the interaction energy (ε=4) is predicted to be <1 kcal/mol and somewhat stronger in higher dielectric.

Computations were also carried out to explore flexibility in the p53-binding site, as it affects the Phe^{86}/Phe^{91} phenyl rings, as a function of substituent (X=H, Cl, Me) in the ligand Trp. For this, starting from the parallel stacked crystal structure conformation (2AXI), additional conformations were generated by rotation of the phenyl ring in Phe^{91} in 10° steps, and from each of the resulting conformations, 10° rotations of the phenyl ring in Phe^{86} were performed, giving a total of 72 conformations. Around 70 % of these conformations produced no steric clashes with the rest of the protein, and for each of these structures binding interaction energy was computed (see appendix 1). This procedure effectively maps out the available space for these two aromatic side chains, and the corresponding interaction energies in the protein environment. From the resulting matrix, the optimal orientation of the two phenyl rings with respect to the ligand X-Trp could be assessed.

In the complexes with cyclic peptides having either X=Cl or X=Me on X-Trp, the conformations with the most favorable ligand-protein interaction energies adopt an arrangement with the two Phe^{86}/Phe^{91} phenyl rings T-stacked (Figure 4-44). However, the cyclic peptide with unsubstituted Trp (X=H) appears to favor a stacked orientation of these rings, which is nevertheless different to that found in the crystal structure with the linear peptide (1YCR). With a linear peptide ligand, where X=H or Me, the phenyl groups adopt a slightly offset tilted stack conformation, whereas when X=Cl the phenyl rings prefer a T-shape conformation.

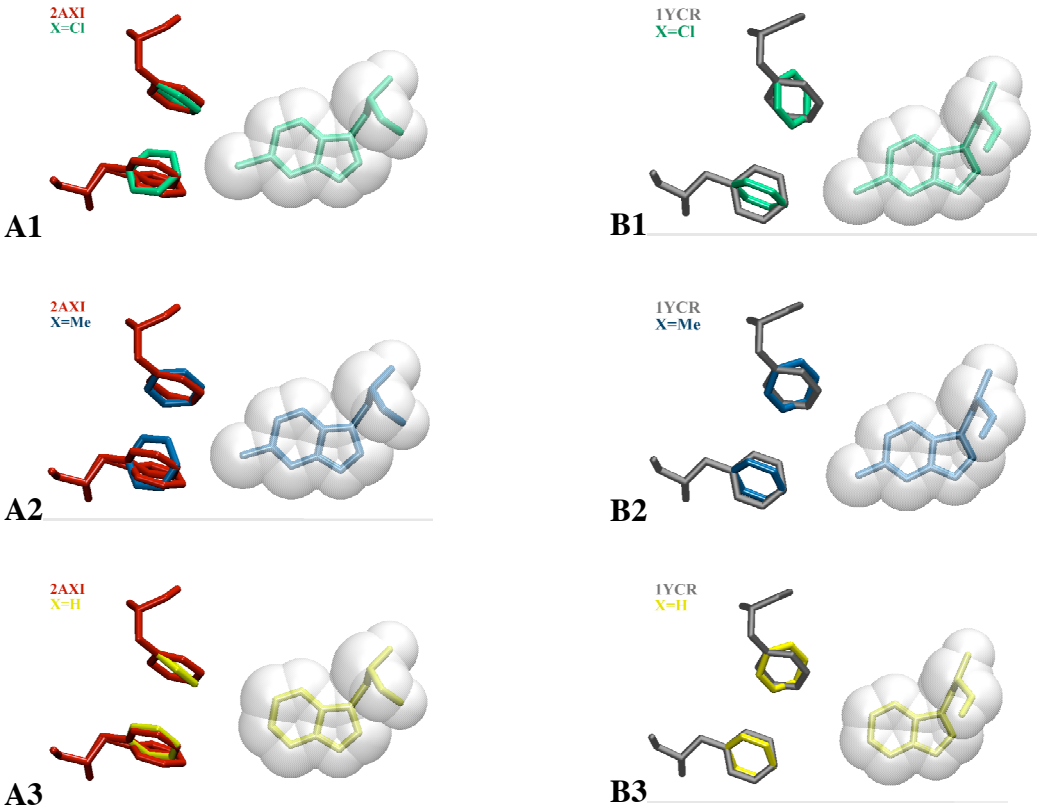


Figure 4-44 Display of X-Trp and Phe^{86}/Phe^{91} from HDM2.

Set **A(1-3)** represent the optimized structures with the cyclic peptide bound, and set **B(1-3)** display the optimized structures with the linear 15-mer peptide. The orientation of the phenyl rings in the initial crystal structures are represented in red for 2AXI and in light yellow for 1YCR. The conformations identified by the conformational search procedure (see text) are shown for X=Cl (green), X=Me (blue), and X=H (yellow).

Several crystal structures of MDM2 receptor bound with an inhibitor was extracted from PDB in addition of the two (1ycr and 2axi) complexes studied in the present work. Table 4-4 describes briefly each complex with the type of orientation between F86 and F91, the substituent used to fill the hydrophobic pocket. The receptor structure remains the same, only the peptide bound to it, is different. Figure 4-45 shows the alignment between 3D crystal structures having a halogen in position 6 on the indole ring. 2GV2 is a small linear 8-er peptide with 6Cl-Trp and 1RV1 is a small peptidomimetic with Bromine substituent.

The crystal structure of the complex with the cyclic peptide shows a shorter distance between the two phenyls than several other crystal structures of the MDM2 receptor bind with different peptide. This short distance could be explained by the beta-hairpin conformation of the peptide, which adopts a double antiparallel strands compare to helix structure from the linear peptides. Whereas all the other crystal structure give the same orientation of both phenyl rings, the complex with the beta-hairpin structure show a stack interaction maybe favored by the shorter distance between the two rings which is about 4Å (optimal distance for a stack conformation is around 3.4Å). In another side the longer distance between the rings favored the orientation of the Hydrogens atoms partially positively charged toward the pi-cloud electrons of the second arene.

PDB ID	X	ORIENTATION F86/F91	DESCRIPTION
1YCR	H	T-shape	MDM2 in complex with a linear peptide (15)
2AXI	Cl	Stack (slide)	MDM2 in complex with a beta-hairpin inhibitor
2GV2	Cl	T-shape	MDM2 in complex with an 8-mer peptide analogue
1T4E	Cl	T-shape	MDM2 in complex with a benzodiazepine
1T4F	H	T-shape	MDM2 in complex with an optimized p53 peptide
1RV1	Br	T-shape	MDM2 in complex with an imidazoline inhibitor

Table 4-4 List crystal structures of HDM2 receptor in complex with a peptide, peptidomimetic or small molecule antogonist to compare the orientation of side chain of Phe86 and Ph91 of the binding site.

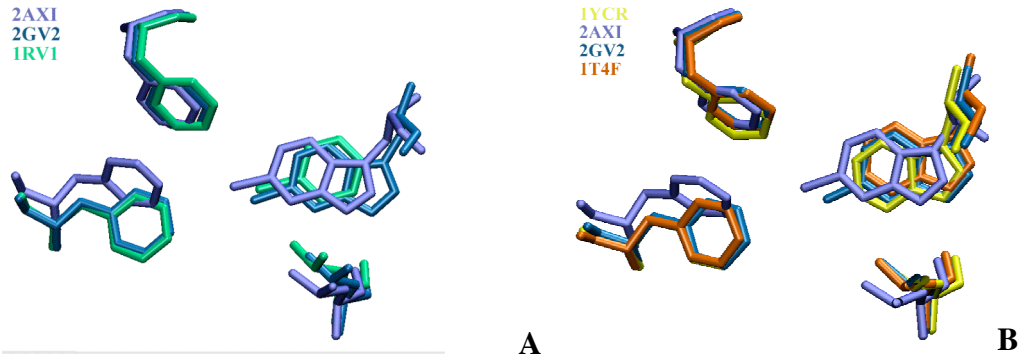


Figure 4-45 Alignment of different inhibitor of HDM2 receptor

A – Structure alignment of HDM2 receptor when it is bound to 8-er peptide analogue (blue), cyclic peptide (violet) in beta-hairpin structure and imidazoline inhibitor (green) with a chlorine atom for the two former peptides and a Bromine atom for the later.

B – Alignment of 4 structures with 2AXI and 2GV2 containing a 6Cl-tryptophan and with 1YCR and 1T4F containing the wildtype tryptophan

4.3.3 Quantify the influence of the distance between W23 of the peptide and F86 and F91 of HDM2 receptor

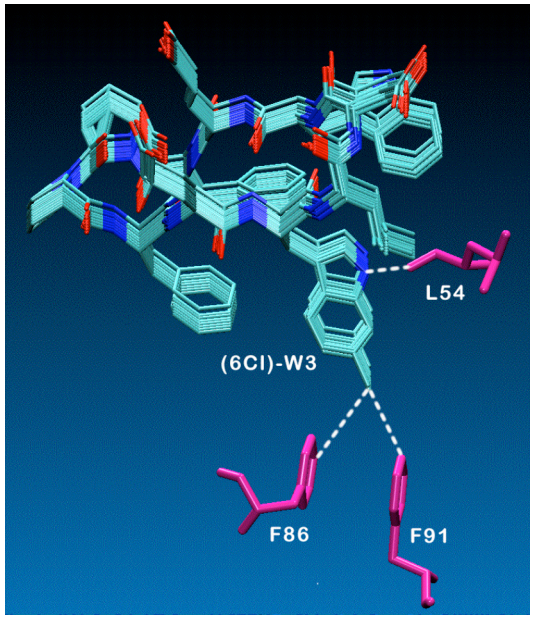


Figure 4-46 Translation of the cyclic peptide away from the binding site of HDM2.

In order to estimate the influence of the distance between the phenyl rings of the residues F86 and F91 from HDM2 receptor and the peptide, slight translations were applied to the complex. The protein position remained unchanged, only the peptide is moved along XY direction away from the binding site. The 3 tables below related the energetic results as well as the relative distances for each new complexes generated. In the same time the distance between the oxygen atom of L54 and the nitrogen atom of the indole ring of the tryptophan is estimated. This distance is important to take in account as it also influence on the overall energetic of the complex as it stabilize the complex. It is interesting to see what happen in term of energy when this interaction is canceled.

The results reported in appendix 2 shows for the 3 complexes a fall in the total binding energy when the two parts are pull apart, especially when the h-bond interaction between the leucine[54] and the tryptophan of interest is eliminated.

4.3.4 Application of the hybrid method to study qualitatively the binding affinity of each complexes

Finally, the calculated electrostatic properties were used to derive binding energies (ΔGbind) for linear and cyclic peptides bound to HDM2, starting from the respective crystal structures (2AXI for the cyclic and 1YCR for the linear peptide), and using the Adaptive Poisson-Boltzmann Solver (APBS) approach. This involved solving the adaptive Poisson-Boltzmann equation for the entire ligand-HDM2 complex, incorporating accurate QM structural and atomic charge data for the peptides. The solvation ($\Delta\Delta G_{solv}$) and coulombic $\Delta\Delta G_{coul}$) contributions to the binding free energy can be computed and used to derive ΔG_{bind} ($\Delta G_{bind} = \Delta\Delta G_{solv} + \Delta\Delta G_{coul}$), where G_{solv} is the solvation free energy, including polar (electrostatic interactions) and nonpolar contributions (dispersion interactions and the energetic costs of creating a cavity in the solvent), and G_{coul} is the coulombic contribution to the binding energy, each of which must be evaluated for the protein ($G_{protein}$), for the ligand alone (G_{ligand}) and for the complex($G_{complex}$).

A breakdown of the contributions to the binding energy is shown in Figure 4-47. The electrostatic contribution shown is the sum of the solvation plus coulombic interactions. The nonpolar term shown represents the dispersion and volume components to the nonpolar solvation forces. As shown, there is a much smaller nonpolar solvation destabilization for binding the cyclic peptide, but also a smaller coulombic stabilization of the complex. The overall trends on the computed total binding energy (ΔG_{bind}) are similar to the experimental observations (ΔH^o), in that the enthalpic contributions in the linear peptides are larger than those in the cyclic peptides.

Table 4-5 summarizes the numerical computational results for the linear and cyclic peptides containing X-Trp, where X=Cl or Me relative to the case where X=H. Here also the qualitative trends in binding energies follow the experimental results in consideration of the substituent effects. A larger preference for X=Cl over X=Me is observed for the cyclic peptide than in the linear case, although both show a preference for the substituent X=Cl.

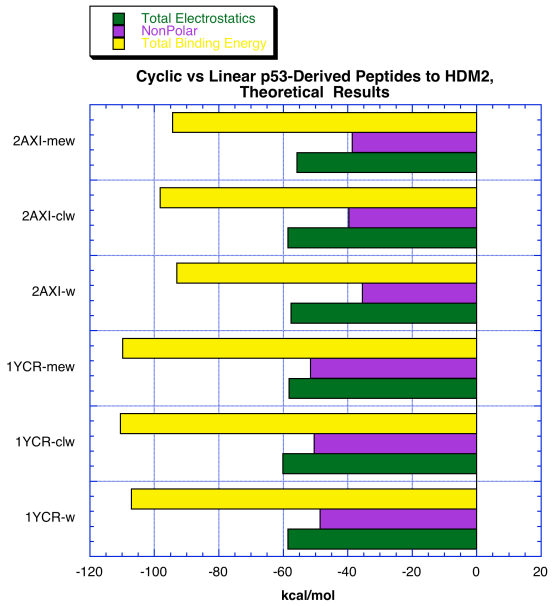


Figure 4-47 Comparison of the calculated thermodynamic signatures for binding of linear (13-15) **vs. cyclic** (6-8) **peptides to rHDM2 determined theoretically**

Table 4-5 Computational results of energetic contributions to the binding energy of the HDM2 receptor and the p53 peptidomimetics 2AXI-(6-8) **and** 1YCR-(13-15)**. Results are expressed relative to the wildtype tryptophan, X=H, in terms of a (ΔΔ) energy difference.**

Complex	$\Delta\Delta G_{elect}$ [1]	$\Delta\Delta G_{nonelect}$ [2]	$\Delta\Delta G_{bind}$ [3]	$\Delta\Delta G_{Expt}$ [4]
2AXI-7(X=Cl)	-0.95	-4.14	-5.09	-0.65
2AXI-8(X=Me)	1.83	-3.12	-1.30	0.17
1YCR-14(X=Cl)	-1.51	-1.90	-3.42	-0.72
1YCR-15(X=Me)	0.34	-3.03	-2.69	-0.66

[1]Electrostatic Contribution (E(gp)+E(ε))

[2]Non-electrostatic Contribution (coulomb+repulsion+cavitation)

[3]Total binding energy

[4]Experimental Binding energy

4.3.5 Comparison with other substituent?

From the computational and experimental studies, a Chlorine substituent favors the binding between the peptidomimetics and the receptor HDM2. Compared to a hydrogen atom that left an empty space between the two molecules, Chlorine seems to fill the space in the hydrophobic cleft of the receptor (Figure 4-48). An investigation was performed using two other halogen substituents. Fluorine atom was choosing as it has a smaller atomic radius than the Chlorine and Bromine atoms for its bigger atomic radius. Table 4-6 reports the results on this study. As expected Fluorine shows a smaller binding energy than Chlorine, certainly due to the free space left in the pocket and a higher electronegativity. The Bromine substituent has also a smaller binding energy than Chlorine, but a larger binding energy compared to fluorine. Bromine fills more space but is less electronegative than Chlorine. From that observation the interactions occurring between the substituent on Trp3 of the peptide and receptor residues require a specific size and electronegativity level to stabilize the complex.

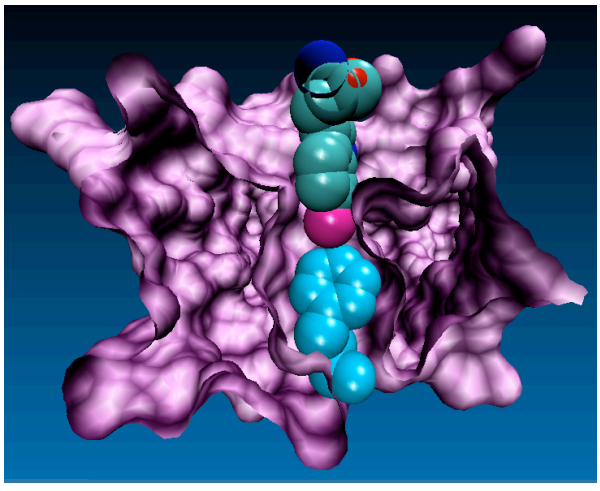


Figure 4-48 6-chlorotryptophan residue embedded into the hydrophobic cleft of the HDM2 receptor and pointing toward the Phe86 colored in blue.

system	ΔΔGsolv	ΔΔGcoul	ΔΔGelec	ΔΔGnp	ΔΔGbind	ΔΔGexp
2axi-6(X=w)	0.00	0.00	0.00	0.00	0.00	0.00
2axi-7(X=clw)	0.62	-1.57	-0.95	-4.14	-5.09	-0.65
2axi-8(X=mew)	0.57	1.25	1.82	-3.12	-1.30	0.17
2axi-(X=fw)	-0.04	-1.84	-1.87	1.22	-0.65	
2axi-(X=brw)	-0.28	-0.04	-0.32	-1.39	-1.71	

Table 4-6 Computational results of energetic contributions to the binding energy of the HDM2 receptor and the p53 peptidomimetics 2AXI-(6-8) 15)**, compare with the results get for complex substituted by fluorine and bromine atom in position 6 of the tryptophane key residue.**

Results are expressed relative to the wildtype tryptophan, X=H, in terms of a (ΔΔ) energy difference in kcal/mol.

4.3.6 Conclusion/Discussion

The binding of HDM2 to helical and hairpin peptide ligands represents an interesting model system to explore the molecular basis of specific, high affinity protein-ligand and protein-protein interactions. Although such interactions are key to many biological processes, it is still extraordinarily difficult to use high-resolution structural information on proteins and the complexes they form, to either predict binding affinities, or to design de novo new molecules able to bind selectively and with high affinity to a chosen target.

We focus on a comparison of the linear (13-15) vs. cyclic (6-8) peptides, where the ITC data reveal a major difference in the thermodynamic signatures of binding to rHDM2. Although the binding of both cyclic and linear peptides is enthalpically driven (negative ΔH^o), the enthalpic effects are 3-4 fold higher with the linear peptides (Table 4-8;Figure 4-50). The ca. 10 kcal/mol higher enthalpic contribution to binding the linear peptides is noteworthy. Conceivably this may arise because the flexibility of the ligand allows adoption of energetically more favorable interactions with the protein than is possible for the more rigid cyclic peptide, and/or due to the formation of new interactions within the p53-derived peptide as it folds (e.g. formation of hydrogen bonds along the α-helix backbone) during binding to HDM2. Note that hydrogen bonds formed within the confines of the protein-ligand binding site should be energetically more important than hydrogen bonds to bulk solvent, due to the change in dielectric. Also, the binding of each cyclic peptide is characterized by a positive $T\Delta S^o$, whereas that of the linear peptides shows a negative $T\Delta S^o$ of similar magnitude. For comparison, when two hydrophobic molecules associate in aqueous solution, with consequent burial of hydrophobic surface, a positive standard entropy of binding ($T\Delta S^o$) is expected, due to release of ordered water molecules from the hydrophobic surfaces of protein and ligand into bulk water.

Examples of negative values of both ΔH^o and $T\Delta S^o$ in the association of molecules in water (with negative ΔG^o) are not exceptional, and have been known for a long time. Although not explainable just in terms of hydrophobic interactions, negative standard entropy of binding may arise due to van der Waals interactions and hydrogen bond formation in the low dielectric environment of the protein interface, and to the requirement for folding of the linear p53-derived peptide into a regular α-helix upon binding. The crystal structures show (qualitatively at least) that the same induced dipole, van der Waals, and H-bonding interactions play a role in stabilizing both the linear- and cyclic-peptide-HDM2 complexes. Hence, it seems likely that the folding of the p53-derived linear peptide must make a major contribution to the 10-12 kcal/mol difference observed in $T\Delta S^o$ values, assuming that the entropic costs of desolvating the linear and cyclic peptides are not greatly different at 298K. It is possibly no coincidence that the differences observed in TS^o values are of the same order as the differences in ΔH^o values, for the linear vs. cyclic peptides, suggesting

entropy-enthalpy compensation during folding and binding of the linear p53-derived peptides.

We now turn to a consideration of the substituent effects on ligand binding to HDM2, in the linear 13-15 and cyclic 6-8 peptides. Although the effects are not as large as those reported earlier for related peptides, the results (

Table 4-7 and Table 4-8, and Figure 4-50) show that adding a Cl-substituent to the indole ring of Trp improves affinity to HDM2 of the linear peptides (14 and 17) by a factor of 5-12 fold, corresponding to a gain in the standard free energy of binding of 1.0-1.5 kcal/mol. Upon adding instead a Me-substituent (15 and 18), the gain in affinity is in the range 4.8-6.1 fold (0.9-1.1 kcal/mol). The ITC data show that the gain in affinity with the Cl-substituent is the result of a more favorable enthalpic contribution (-1.7 kcal/mol), offset by a more positive entropic contribution (+0.7 kcal/mol). In the cyclic peptides, however, adding the chloro-substituent (11 and 7) improves affinity only by a smaller amount (2-3 fold; 0.4-0.6 kcal/mol), and the change is driven entirely by an entropic effect. In contrast, adding instead a methyl substituent (12 and 8) worsens affinity by a factor of 0.4-0.7 (0.2-0.5 kcal/mol), which seems to be mostly due to a decrease of the enthalpic contribution.

At first sight, the p53-HDM2 interaction involves mainly conserved hydrophobic contacts, including hydrophobic and aromatic groups in p53 and in the deep hydrophobic pockets on HDM2. We have used hybrid QM/PB techniques here in an attempt to gain greater insight into the electronic nature of these interactions. One interaction of special interest is the hydrogen bond between the indole NH of the X-Trp residue in the linear and cyclic peptide ligands, and the backbone carbonyl oxygen of Leu[54]. The energetics of this H-bond were investigated using not only the available crystal structures (i.e. as observed in the crystalline state at low temperature), but also using optimized interaction geometries. Each crystal structure represents just one of perhaps many possible interaction geometries, which can arise due to protein fluctuations in solution at room temperature. It is, therefore, helpful to explore the extent to which interaction energies change, as the geometries of the interacting groups are individually optimized.

The geometry optimized structures show the hydrogen bond to be more elongated and directed along the hydrogen-to-oxygen lone pair vector, than is observed in the crystal structure (Figure 4-41d). The calculated structures, consisting only of the X-Trp-Leu54 fragment in the absence of the additional surrounding residues, enables the methyl side chain hydrogens of Leu54 to position themselves over the aromatic ring, with 3 H-p interactions (for X=Cl), or 2 H-p interactions (for X=CH3). The resulting, more compact structures, have a much smaller dipole moment than in the crystal structure, where the methyl groups interact with the surrounding protein. In the actual protein environment, hydrogens from surrounding isoleucines (Ile99 and Ile103), and Leu4 from the peptide itself, point into the face of the tryptophan (Figure 4-41e).

The calculated strength of the optimized hydrogen bond to the indole NH ranges from 10.6-11.4 kcal/mol in the gas phase, with the case X=H having the strongest interaction. In a lower protein-like dielectric environment (ε= 4), the H-bond values are reduced to 5.8-6.5 kcal/mol. Using the conformation present in the crystal structure, the computed H-bond energies in a protein-like dieletric are still 2.1-2.5 kcal/mol, now with X=Cl showing the strongest interaction. However, it seems possible that fluctuations in the protein in solution near room temperature may well occur, which recover more of the intrinsic H-bonding energy described by the optimized geometries. We conclude, therefore, that this anchoring hydrogen bond represents an important source of binding energy, which contributes considerably to affinity and most likely also to the specificity of the interaction. Interestingly, theory predicts that introduction of the chlorine substituent in the cyclic peptides will strengthen the H-bond in the crystal structure geometry.

A second interaction site of special interest surrounds the indole nucleus of the X-Trp residues in 6-8 and 13-15. Concerning first the flexibility of the X-Trp-binding site, the calculations suggest that there is sufficient flexibility in the binding site in the crystal structure to allow the Phe^{86} and Phe^{91} aromatic groups to adopt a range of $\chi 2$ torsion angles. The magnitude of the interaction energies between these side chains and the X-Trp is dependent upon the precise interaction geometries, and in particular upon the dielectric in the binding pocket. The calculations suggest that small but significant stabilizing interactions occur between the polarizable chlorine substituent and the partial positive charge of the phenyl hydrogen atoms. In the case of the methyl substitution, computations suggest small but significant stabilizing vdW interactions between hydrogens of the Me group and hydrogens of the Phe^{86} and Phe^{91} side chains. However, interactions of a similar magnitude are also seen for the unsubstituted case, where X=H. With the methods available at present, the calculations are not sufficiently accurate to provide a reliable ranking of these individual small (<1 kcal/mol) interaction energies.

Advances in computational methods promise to facilitate greatly the complex process of structure-based protein ligand design. However, it remains a major technical challenge to accurately predict ligand-receptor binding affinities using theoretical methods. The available X-ray structural and experimental thermodynamic binding data make these ligand-protein interactions an interesting model system to examine using current theoretical methods. One approach to calculate binding energies, and in particular binding energy differences, is to evaluate the electrostatic properties of the system, including coulombic and solvation energy contributions, by solving the Poisson-Boltzmann equation. The calculations used, however, do not fully account for the changes in entropy that occur upon binding, which are very complex in this system (vide infra), and so the calculated binding energies (ΔG_{bind}) are not equivalent to the standard free energy of binding (ΔG^{o}). Also, a direct comparison of computational binding energies (ΔG_{bind}) to experimental ΔH^{o} and $T\Delta S^{o}$ values is not possible. However, the values of both ΔG_{bind} and ΔH^{o} will certainly be influenced by changes in the electrostatic properties of the bound ligand, as the structure is varied (e.g. linear vs cyclic ligand, chloro- vs methyl-substituent (X)), and so correlations between binding energy differences predicted by theory and those observed experimentally might be expected.

The calculated total binding energy (ΔG_{bind}) (Figure 4-47, yellow) includes both electrostatic interactions (green) arising from coulombic and polar solvation contributions to binding, as well as nonpolar solvation effects (purple) arising from cavitation and dispersion terms. The results shown in Figure 4-47 clearly show stronger binding energy in the linear (13a) vs. the cyclic peptide (7), which is indeed reflected in the larger ΔH^o values determined for the linear peptides by ITC (Table 4-8). This is shown to arise computationally from a difference in the nonpolar contributions to solvation in the two systems, since the electrostatic contributions (green) are not largely different.

Considering now the differences caused by changes to the substituent X in the X-Trp residue; the small increase in binding (ΔGbind) for each ligand with X=Cl over X=Me (compared to X=H) in the linear peptides, detected experimentally in the ΔH^o (and ΔG^o) values (Table 4-8), is reproduced in the computational results. The results shown relative to the case where X=H illustrate this latter point more clearly (Figure 4-51). In addition, relative to the case where X=H, the peptides with X=Me in both linear (1YCR) and cyclic (2AXI) forms, show decreased electrostatic interactions, although to a greater degree in the linear form. **Figure 4-49** illustrates the calculated changes in molecular electrostatic potential (MEP) profiles for the complexes with linear and cyclic ligands, for the case X=Me (left side) relative to X=H, and for the case X= Cl (right side), each relative to the case X=H. The most dramatic effects are observed in the cyclic peptide, which clearly show the change in electrostatic nature of the peptide when the substituent is changed from X=H to X=Me (additional positive charge indicated by blue), and from X=H to X=Cl (additional negative charge indicated by red). On the other hand, in the linear peptide, while the surface of the peptide is very slightly more negative with the addition of chlorine (X=Cl, left side) and blue with the addition of the methyl (X=Me, right side), the response is much less than for the cyclic peptide.

The ability of this theoretical approach to correctly reflect key thermodynamic features of the binding reactions observed experimentally for this series of peptide ligands, and provide insights into the electronic nature of the interactions, is notable. Such peptide-protein interactions represent interesting targets for the further development of the peptidomimetic approach to ligand design, and of computational

methods for the analysis of protein-ligand interactions, especially for the difficult case of discovering protein-protein interaction inhibitors.

LIGAND (SUBSTITUENT)	K_D (μM)	NORMALIZED $1/K_D$ (ΔG KCAL/MOL)
10 (X = H)	0.12 ± 0.01	1
11 (X = Cl)	0.065 ± 0.02	1.85 (0.36)
12 (X = Me)	0.27 ± 0.07	0.43 (-0.50)
16	0.67 ± 0.07	1
17 (+Cl)	0.055 ± 0.005	12.2 (1.48)
18 (+Me)	0.11 ± 0.01	6.1 (1.07)

Table 4-7 Dissociation constants (KD) determined by SPR for the interactions of 10-12 and 16-18 with rHDM2 at 298K. Measurements were performed in triplicate and the mean and deviations are shown.

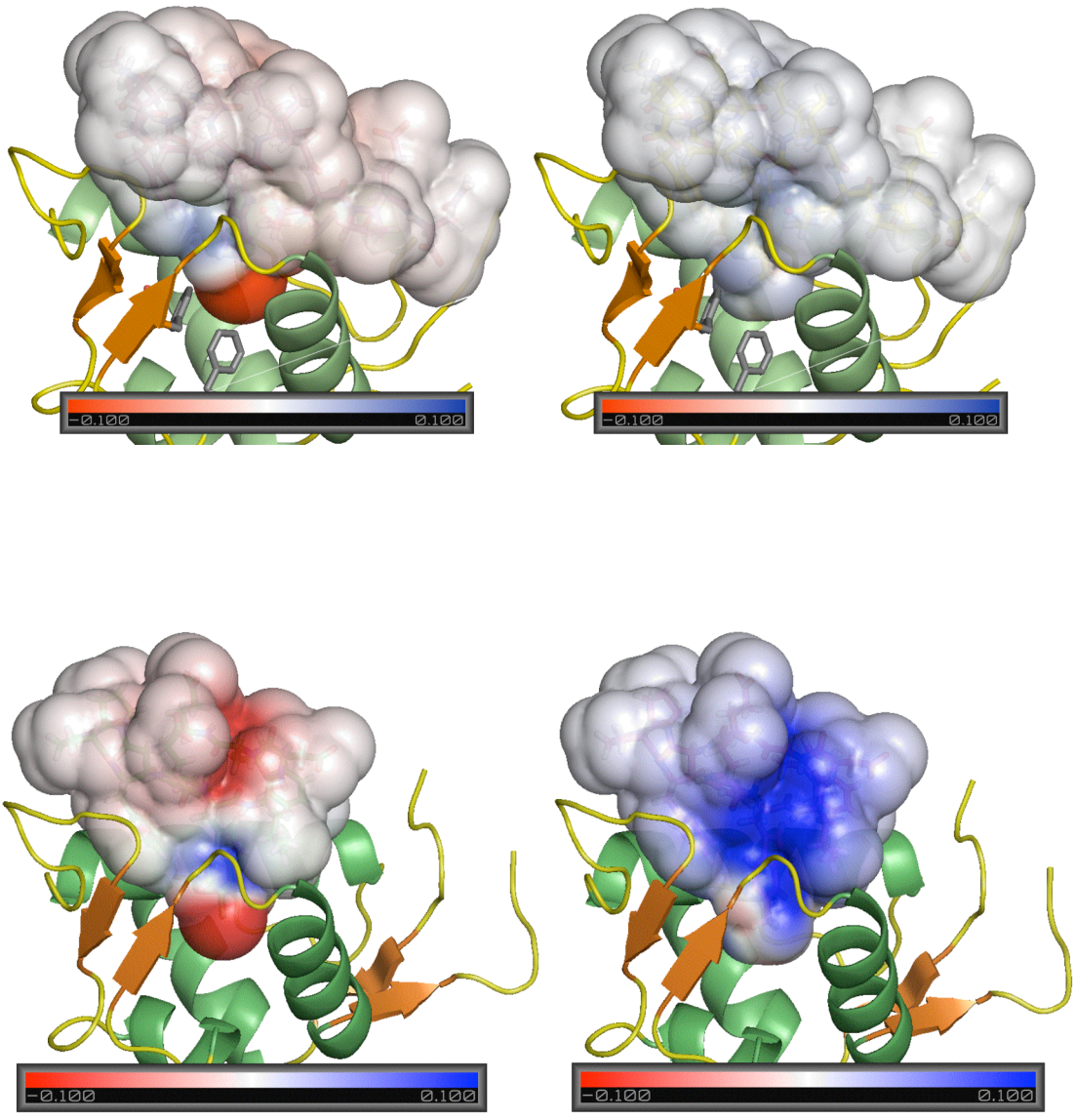
Ligand (substituent)	ΔH^o (Kcal/mol)	$T\Delta S$ (kcal/mol)	Stoichio-metry	K_D (μM)	Normalized $1/K_D$ (ΔG kcal/mol)
6 (H)	-4.6±0.1	+5.1±0.2	0.94±0.03	0.073±0.005	1.0
7 (Cl)	-4.6±0.1	+5.8±0.4	0.92±0.02	0.025±0.002	2.92 (0.63)
8 (Me)	-3.9±0.1	+5.7±0.2	0.94±0.1	0.10±0.02	0.73 (-0.19)
13 (H)	-14.1±0.7	-5.4±0.7	0.91±0.1	0.43±0.03	1.0
14 (Cl)	-15.8±0.6	-6.1±0.5	1.01±0.1	0.08±0.01	5.38 (1.00)
15 (Me)	-14.7±0.6	-5.1±0.7	1.00±0.09	0.09±0.06	4.78 (0.93)

Table 4-8 Ligand binding to rHDM2 measured by isothermal titration calorimetry at 298K.

Measurements were performed at least in duplicate and the mean and deviations are shown.

Figure 4-49 Display of HDM2-ligand complexes.

Set **top** represent the molecular electrostatic potential map difference between X=Cl and X=H (left side) and between X=Me and X=H (right side) in the linear peptide bound (as in 1YCR), and **bottom** display the molecular electrostatic potential map difference between X=Cl and X=H (left side) and between X=Me and X=H (right side) in the cyclic peptide bound (as in 2AXI).



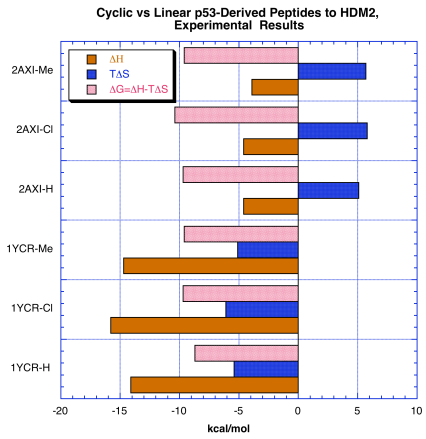


Figure 4-50 Comparison of the experimental thermodynamic signatures of binding to rHDM2 for the linear (**13**-**15**) vs. cyclic (**6**-**8**) peptides

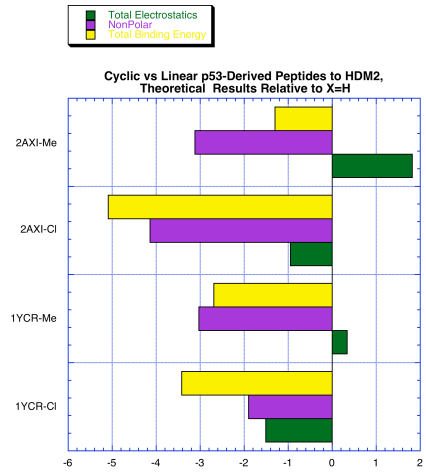


Figure 4-51 Comparison of the calculated thermodynamic signatures for binding of linear (13-15) vs. Cyclic (6-8) peptides to rHDM2 determined theoretically

5 Implemented a loop between GAMESS-APBS in order to include the environment charges around the ligand in the Fock operator Self-consistently.

5.1 Introduction and motivation

Today the most common Continuum Electrostatic model used to study macromolecular electrostatics is based on Poisson–Boltzmann Equation (PBE). An adaptive, multilevel finite element method was developed and incorporated in APBS to solve numerically the non-linear PBE in order to estimate the electrostatic contribution to the free energy of solvation of molecules immersed in aqueous environment. This theory describes the solvent as a continuous high dielectric and ionic concentration with uniform density. The solute, the molecules are representing classically as a collection of charges and positions. Using APBS solves PBE and the electrostatic potential is estimated at each point of the 3D grid defined to perform the calculation. In the hybrid method presented here, one of the goals is to get accurate properties for the ligand. In order to compute accurate atomic charges of the ligand, this later should be replaced in the same environment create by the protein. It is important to get the effect of the electric field produced by the protein on the charge distribution of the ligand. The present work has the goal of creating a bridge between APBS [122] and the Quantum chemistry program GAMESS (US) [55]. Integrating into GAMESS the electrostatic potential grid calculated by APBS is a way to get the polarization from the protein onto the ligand. The protein generates an electric field, which acts on the ligand when it interacts with the protein. This perturbation can be introduced into the Hamiltonian by using first principles of Quantum Mechanical theory. Due to approximation in Hartree–Fock equation to reduce the many-body problems to one electron and to treat the electron-electron repulsion with an average field, the external field can be added directly to the electronic Hamiltonian. Self-Consistent Field Reaction procedure is used to reach convergence. The partial atomic charges of the ligand are then putting back into PBE to estimate more accurately the affinity between the ligand and the protein. This additional step will improve the charge distribution of the ligand by taking in account the surrounding in which the ligand is embedded.

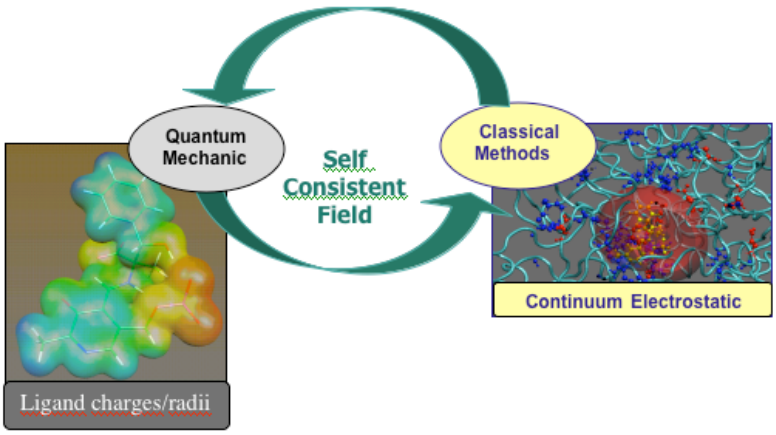


Figure 5-1 Overview of the Self-consistent loop between GAMESS and APBS

5.2 Self-Consistent method implemented in GAMESS

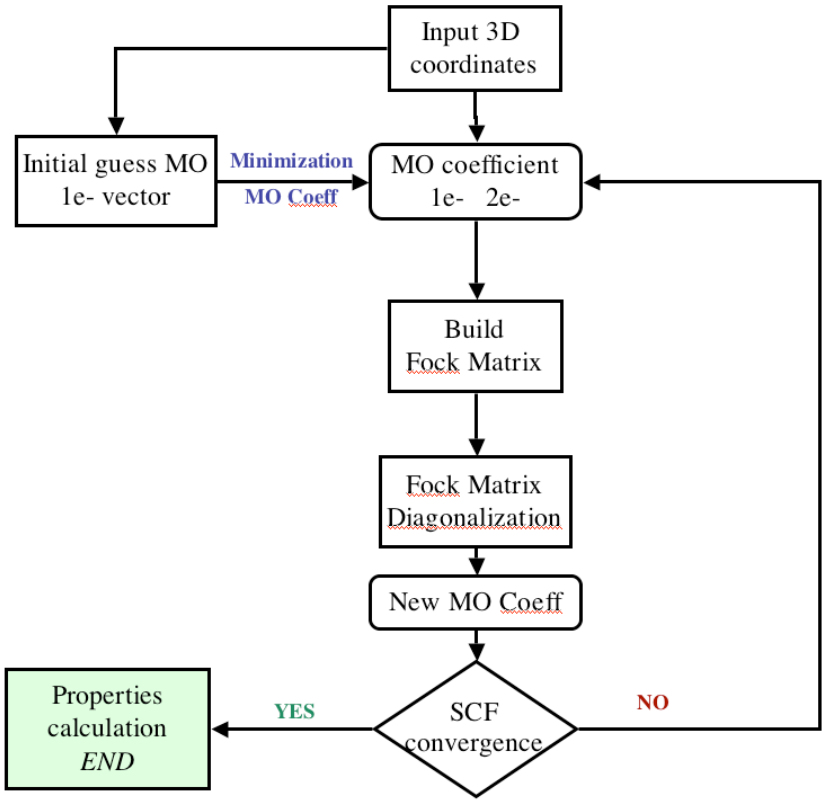


Figure 5-2 SCF methods in GAMESS

The external field generated by APBS is integrated into the One-electron effective Hamiltonian of GAMESS. The Hamiltonian carries the energy components due to interaction between electrons, nuclei and nuclei/electrons. It is an operator acting on the wavefunction to solve the Schrödinger equation and get the energy of the system. The Hamiltonian is defined as a sum of unperturbed Hamiltonian and potential energy:

$$\boxed{H = H^0 + \Delta V}$$

5.3 Implementation steps

5.3.1 Reading of the DX file from APBS into GAMESS

Along electrostatic calculations, files can be generated to write the electrostatic potential estimated by solving the non-linear PBE on each point of the grid used to perform the calculation. The format file use by APBS to write the electrostatic potential is a DX format. Few lines at the beginning give the information regarding the grid used to solve the PBE. Then the dimensionless electrostatic potential values are written by 4 values per line. Each value corresponds to an electrostatic potential estimated at specific point of the grid.

The DX file begins by giving all the parameters describing the grid: size (along x, y, z axis), grid spacing, number of grid point. Following those lines, the electrostatic potential is listing. The format described the electrostatic by beginning from the lower left corner of the grid and by varying the z position faster, followed by the y, and then followed by x. Therefore the file contains all the information in order to get the Cartesian coordinates corresponding to grid point with an electrostatic value associated. The grid unit is converted into Cartesian coordinates thanks to the information related to the grid.

The source code implemented to read dx file in games is giving in Appendix 3.

5.3.2 Copy dxfile in memory for a games run

The dx file describing the electrostatic potential of the system has to be stored in memory along the computation. The dx file is copied in a Direct Access File (DAF), which is special file used in GAMESS to put in memory external data required for the job. The dx file has to be in the same directory as the input file and differ by this later only by the extension file (dx instead of inp). The file is loading temporarily in memory under the extension F339 thank to the environment variable DXNAME define in rungms script.

5.3.3 Integration of new functionality in GAMESS

GAMESS is a quantum chemistry package written in fortran with C module to interface Gamess with the system. To interact with gamess, the user writes an input file with a special format. This input file constitutes by modules arranged in dollar groups. Each dollar group gathers an ensemble of same scientific meaning. In order to include the new functionality in GAMESS to create the bridge with APBS a new dollar group was implemented: $APBS. To make gamess read the new group, a new program was added in the source code where several routines were implemented and will be described along this chapter. This dollar group has several dependencies with other dollar group existing in gamess. It required the use of the $ELPOT group in order to make an electrostatic study and to specify which charges the program will use. For now the charges use are CHELPG. This require WHERE=PDC and the $PDC group need to turn on with PTSEL keyword set to CHELPG.

5.3.3.1 APBSIN

This routine allows the program to read what users write in the input file. It is in general the first subroutine written in a source file code.

The routine is called from the **inputa.src** file. It checks the validity of the keywords, whether there is no error in the spelling of one keyword, but also compare the type of data required for the specific keyword.

Four keywords compose the new dollar group $APBS group:

- **LAPBS**: integer telling the kind of computation proceed with APBS
    - 0 - Skip the property (default)
    - 1 - Binding Energy
- **NAPROT**: integer specifying the protein pqr file
    - 0 - No protein PQR file associated with the computation (default)
    - 1 - A protein PQR file associated with the computation. The user need to name his file as the same as his input file, with the prefix 'prot': xxx.inp and prot_xxx.pqr
- **NACOMP**: integer specifying the complex pqr file
    - 0 - Do not punch complex file (default)
    - 1 - Punch out a complex file with the new charges of the ligand
- **NALIG**: integer to punch a ligand PQR file
    - 0 - Do not punch ligand file
    - 1 - Punch a ligand PQR file (default)

5.3.3.1.1 LOGICAL IAPBS, UAPBS

IAPBS and UAPBS are logical variables. IAPBS checks if the keyword $APBS is turning on and UAPBS checks if the first loop was completed, and then turn it to true to get the energy with the influence of the field.

5.3.3.1.2 APBSDX

APBSDX is the routine to read the dx file. Along the reading, the Cartesian coordinates of each point of the grid are stored in a 3D array called CORAPBS. The electrostatic potential values at each point of the grid are stored in a 1D array called POTAPBS. This array is dimensioned by the value corresponding to the number of points in the grid, NPGRD.

According to the Boolean variable UAPBS, a first loop is performed in gas phase to estimate the energy, GAPBS. A second loop is performed by including the electrostatic potential grid into the one-electron integral. This give the energy with the electrostatic potential surrounding, EAPBS. And the energy difference between is estimated with DAPBS

A new common block was added into the routine: APBSFD

```
COMMON /APBSFD/ GPAPBS,EAPBS,DAPBS,POTAPBS(NPGRD),

     *          CORAPBS(3,NPGRD)
```

6 Conclusion

In the past decade, there have been significant contributions in the development of algorithmic procedures and associated scoring functions used for computational investigation of small molecules interacting inside a protein. The main goal of such investigations is to provide structure and mechanistic information for better understanding this interaction process. Such computational capability has had a large impact in the understanding of ligand-protein and protein-protein binding kinetics and overall development of new drugs.

The goal of the computational approach described in this thesis, was to create hybrid computational methodology for study of ligand-protein interactions based on quantum mechanical techniques and classical techniques for modeling biomolecular interactions and solvation through solution of the PB equation, thereby providing accurate technology for understanding conformation of small molecules in active sites, with a corresponding accurate scoring function to predict the affinity between a ligand and its target receptor.

The resulting hybrid method was successfully applied on two key biomolecular systems, where the additional QM computation enables one to explore more detail with greater accuracy of the interaction of fragments in the binding pocket that are involved in the binding process. Application 1 involved a research collaboration with the experimental group of D. Hilvert (ETH), to show that an engineered alr catalyzes an unprecedented retroaldol reaction of α-substituted β-phenylserines with high stereospecificity at the α-carbon but low stereochemical control at the β-carbon. Not only is this reaction catalyzed some 10^4-fold more efficiently than by the cofactor in solution, but the hybrid computational model provided a basis for further optimization of the alrY265A active site via site directed mutagenesis or directed evolution. The goal is a practical catalyst for the enantioselective synthesis of novel amino acids and derivatives. In this study, 4 diastereomers were considered in the active site, and showed that, in accord with the experimental findings, the modified binding pocket readily accommodates only two of them, both exploiting very similar interactions with the binding pocket and the cofactor. This work has been published in Angewante. Chemie. Int. Engl, **2006**, 45, 6824 [153].

Application 2 involved a research collaboration with the experimental group of J. Robinson (UZH) involving the interaction of p53 peptidomimetics with the HDM2 protein. In this investigation, both the experimental thermodynamic signatures as well as the computational electronic nature of interactions experienced by a series of linear p53-derived peptides and cyclic β-hairpin peptidomimetics in binding to HDM2, was studied. Computations were used to elucidate the electronic nature of key interactions between each ligand and HDM2, as well as global binding energies. The calculations characterize an energetically important hydrogen bond from a key Trp indole NH in each ligand to a peptide backbone carbonyl in HDM2, as well as stabilizing interactions between aromatic H-atoms in key Phe groups in the surrouding residues, as well as a 6-substituent in the indole of the key Trp residue in each ligand. In particular, the ligand with a chlorine substitution on this Trp residue was seen to increase the affinity of the peptide greatly, and both experiment and computation was used to understand this. This work has been submitted to J. Am. Chem. Soc. for consideration.

The complex nature of the associated computational methodology presented here was facilitated by the development of new tools involving grid infrastructure. The framework GEMSTONE was motivated by supporting the execution of the different steps involved in the hybrid method, with integration of computational as well as analysis software to provide a complete working environment for this particular type of research.

Further work should focus on application of the hybrid method on a large scale set of applications and ligand-protein systems, by using the more automated grid infrastructure combined with a workflow engine. Along this line, preliminary work has been done involving grid software (Nimrod/G) to perform automatically the docking and evaluation of electronic interaction of, for example, the PLP bound substrate on the aldolase. The adaptation of this experiment to a workflow engine called KEPLER should motivate further development to manage the behavior of the workflow's execution. Such capabilities will integrate the middleware tool NIMROD into the workflow engine, KEPLER, to distribute the computational procedure over grid resources. Such technology will enable much more expansive theoretical investigation over more of the parameter space that is possible by hand.

The limit of the method present here is the lack of dynamical motion considered in the large biosystem. One improvement on the methodology is the development of a fully integrative method, bringing the entire ligand environment into the quantum mechanical self-consistent field loop, thereby enabling one to consider all motion of the ligand in the field of the protein and the environment at once. A second improvement would involve self-consistent relaxation of the entire system via an integrated classical method such as provided by MD.

7 References

1. Lane, D.P., *Cancer - P53, Guardian of the Genome*. Nature, 1992. **358**(6381): p. 15-16.
2. Chen, J., J. Lin, and A.J. Levine, *Regulation of transcription functions of the p53 tumor suppressor by the mdm-2 oncogene*, in *Mol Med*. 1995. p. 142-52.
3. Momand, J., et al., *The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation*, in *Cell*. 1992. p. 1237-45.
4. Hu, S.H., *Efficient enzymatic process for the production of (2S)-4,4-difluoro-3,3-dimethyl-N-Boc-proline, a key intermediate in the synthesis of HIV protease inhibitors*. Organic Process Research & Development, 2006. **10**(3): p. 650-654.
5. Griffiths, J.S., et al., *A bacterial selection for the directed evolution of pyruvate aldolases*. Bioorganic & Medicinal Chemistry, 2004. **12**(15): p. 4067-4074.
6. Seebeck, F.P., *Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation*. Journal of the American Chemical Society, 2003. **125**(34): p. 10158-10159.
7. Dalby, P.A., *Engineering Enzymes for Biocatalysis*, in *recent Patents on biotechnology 2007*. 2007. p. 9.
8. McCammon, J.A., *Theory of biomolecular recognition*, in *Curr Opin Struc Biol*. 1998. p. 245-249.
9. Gellman, S.H., *Introduction: Molecular recognition*, in *Chem Rev*. 1997. p. 1231-1232.
10. GILSON, M.K. and B.H. HONIG, *CALCULATION OF ELECTROSTATIC POTENTIALS IN AN ENZYME ACTIVE-SITE*, in *Nature*. 1987. p. 84-86.
11. McCammon, J.A., *COMPUTER-AIDED MOLECULAR DESIGN*, in *Science*. 1987. p. 486-491.
12. Gilson, M.K., *Theory of electrostatic interactions in macromolecules*, in *Curr Opin Struc Biol*. 1995.
13. Ma, C.S., et al., *Binding of aminoglycoside antibiotics to the small ribosomal subunit: A continuum electrostatics investigation*, in *Journal of the American Chemical Society*. 2002. p. 1438-1442.
14. TOMASI, J., *Thirty years of continuum solvation chemistry: a review, and prospects for the near future*, in *Theor Chem Acc*. 2004. p. 184-203.
15. Feig, M. and C.L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, in *Curr Opin Struc Biol*. 2004. p. 217-224.
16. Warren, G.L., et al., *A critical assessment of docking programs and scoring functions*, in *J Med Chem*. 2006. p. 5912-5931.

17. Huang, N., B.K. Shoichet, and J.J. Irwin, *Benchmarking sets for molecular docking*, in *J Med Chem*. 2006. p. 6789-6801.

18. Perutz, M.F., *Electrostatic effects in proteins*, in *Science*. 1978. p. 1187-91.

19. Debye, P. and E. Huckel, *The theory of electrolytes I. The lowering of the freezing point and related occurrences*, in *Phys Z*. 1923. p. 185-206.

20. WARWICKER, J. and H.C. WATSON, *CALCULATION OF THE ELECTRIC-POTENTIAL IN THE ACTIVE-SITE CLEFT DUE TO ALPHA-HELIX DIPOLES*, in *J Mol Biol*. 1982. p. 671-679.

21. Baker, N.A., *Improving implicit solvent simulations: a Poisson-centric view*, in *Curr Opin Struc Biol*. 2005. p. 137-143.

22. Sharp, K.A., R. FINE, and B. Honig, *COMPUTER-SIMULATIONS OF THE DIFFUSION OF A SUBSTRATE TO AN ACTIVE-SITE OF AN ENZYME*, in *Science*. 1987. p. 1460-1463.

23. JEANCHARLES, A., et al., *ELECTROSTATIC CONTRIBUTIONS TO SOLVATION ENERGIES - COMPARISON OF FREE-ENERGY PERTURBATION AND CONTINUUM CALCULATIONS*, in *Journal of the American Chemical Society*. 1991. p. 1454-1455.

24. Nielsen, J.E. and J.A. Mccammon, *Calculating pKa values in enzyme active sites*, in *Protein Sci*. 2003. p. 1894-1901.

25. Oberoi, H. and N.M. Allewell, *Multigrid solution of the nonlinear Poisson-Boltzmann equation and calculation of titration curves*, in *Biophysical Journal*. 1993.

26. BURKERT, U. and N.o.r.m.a.n. L. Allinger, *Molecular Mechanics*. 1982. p. 340.

27. KUNTZ, I.D., et al., *A GEOMETRIC APPROACH TO MACROMOLECULE-LIGAND INTERACTIONS*, in *J Mol Biol*. 1982. p. 269-288.

28. Gohlke, H., M. Hendlich, and G. Klebe, *Knowledge-based scoring function to predict protein-ligand interactions*, in *J Mol Biol*. 2000. p. 337-356.

29. Eldridge, M.D., et al., *Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the …*, in *J Comput Aid Mol Des*. 1997.

30. Solomonides, T.o.n.y. and R.i.c.h.a.r.d. McClatchey, *From Grid to Healthgrid: Proceedings of Healthgrid 2005 - Page 293*. 2005. 324.

31. Bonetto, P., M. Guarracino, and F. Inguglia, *Integrating medical Imaging into a Grid based computing infrastructure*, in *Lect Notes Comput Sc*. 2004. p. 505-514.

32. Buyya, R., et al., *Neuroscience instrumentation and distributed analysis of brain activity data: a case for eScience on global Grids*, in *Concurr Comp-Pract E*. 2005. p. 1783-1798.

33. Chrisochoides, N., et al., *Imaging and visual analysis---Toward real-time image guided neurosurgery using distributed and grid …*, in *Proceedings of the 2006 ACM/IEEE conference on …*. 2006.

34. Dimaio, S.P., et al., *Image-guided neurosurgery at Brigham and Women's Hospital*. Engineering in Medicine and Biology Magazine, IEEE, 2006. **25**(5): p. 67 - 73.

35. Devadithya, T., et al., *On-demand high performance computing: image guided neuro-surgery feasibility study*, in *Parallel and Distributed Systems, 2006. ICPADS 2006. 12th International Conference on*. 2006. p. 6.

36. Rossi, L., *The Large Hadron Collider and the Role of Superconductivity in One of the Largest Scientific Enterprises*, in *Applied Superconductivity, IEEE Transactions on*. 2007. p. 1005 - 1014.

37. Gorder, P.F., *Physics Experiment Could Spawn Permanent Computing Grid*, in *Computing in Science & Engineering*. 2007. p. 5 - 9.

38. Foster, I. and C. Kesselman, *The Globus project: a status report*, in *Heterogeneous Computing Workshop, 1998. (HCW 98) Proceedings. 1998 Seventh*. 1998. p. 4 - 18.

39. Foster, I. and C. Kesselman, *The Globus toolkit*, in *The grid: blueprint for a new computing infrastructure table ….* 1998.

40. Sudholt, W., et al., *Applying grid computing to the parameter sweep of a group difference pseudopotential*. Computational Science - Iccs 2004, Pt 1, Proceedings, 2004. **3036**: p. 148-155.

41. Altintas, I., et al., *A framework for the design and reuse of grid workflows*, in *Lect Notes Comput Sc*. 2004. p. 120-133.

42. Mock, S., et al., *The computational chemistry prototyping environment*, in *P Ieee*. 2005.

43. Mock, S., et al., *GEMSTONE: GRID ENABLED MOLECULAR SCIENCE THROUGH ONLINE NETWORKED ENVIRONMENTS*, in *Grid Computing in Life Sciences*. 2006.

44. Baldridge, K.K., et al., *The computational chemistry prototyping environment*. Proceedings of the Ieee, 2005. **93**(3): p. 510-521.

45. Abramson, D., et al., *Nimrod: a tool for performing parametrised simulations using distributed workstations*, in *High Performance Distributed Computing, 1995., Proceedings of the Fourth IEEE International Symposium on*. 1995. p. 112-121.

46. Abramson, D., et al., *The Laboratory Bench: Distributed Computing for Parametised Simulations*, in *1994 Parallel Computing and Transputers Conference*. 1994.

47. Foster, I., *The Grid: A New Infrastructure for 21st Century Science*, in *Physics Today*. 2002.

48. Foster, I. and C. Kesselman, *The grid: blueprint for a new computing infrastructure*. 2004. p. 748.

49. Amoreira, C. and K.K. Baldridge, *Molecular Mechanics*, in *In McGraw Hill Encyclopedia of Science and Technolgoy*, B. 10th Ed., D., Ed., McGraw Hill Professional, Editor. 2006: New York.

50. Lennard-Jones, J.E., *Cohesion*, in *P Phys Soc*. 1931. p. 461-482.

51. Jorgensen, W.L., D.S. Maxwell, and J. TiradoRives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*, in *Journal of the American Chemical Society*. 1996. p. 11225-11236.

52. Sitkoff, D., K.A. Sharp, and B. Honig, *Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models*, in *Journal of Physical Chemistry*. 1994. p. 1978-1988.

53. ROOTHAAN, C.C.J., *NEW DEVELOPMENTS IN MOLECULAR ORBITAL THEORY*, in *Rev Mod Phys*. 1951. p. 69-89.

54. POPLE, J.A. and R.K. NESBET, *SELF-CONSISTENT ORBITALS FOR RADICALS*, in *Journal of Chemical Physics*. 1954. p. 571-572.

55. Schmidt, M.W., et al., *General Atomic and Molecular Electronic-Structure System*. Journal of Computational Chemistry, 1993. **14**(11): p. 1347-1363.

56. Moller, C. and M.S. Plesset, *Note on an approximation treatment for many-electron systems*, in *Phys Rev*. 1934. p. 0618-0622.

57. Zhao, Y. and D.G. Truhlar, *The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals*, in *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*. 2008. p. 215-241.

58. Mulliken, R.S., *Electronic Population Analysis on Lcao-Mo Molecular Wave Functions .1*. Journal of Chemical Physics, 1955. **23**(10): p. 1833-1840.

59. F. W. Bader, R.i.c.h.a.r.d., *Atoms in Molecules: A Quantum Theory*. 1994. p. 458.

60. Reed, A.E., *Natural-Population Analysis*. Journal of Chemical Physics, 1985. **83**(2): p. 735-746.

61. BAYLY, C.I., et al., *A WELL-BEHAVED ELECTROSTATIC POTENTIAL BASED METHOD USING CHARGE RESTRAINTS FOR DERIVING ATOMIC CHARGES - THE RESP MODEL*, in *Journal of Physical Chemistry*. 1993. p. 10269-10280.

62. Breneman, C.M., *Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis*. Journal of Computational Chemistry, 1990. **11**(3): p. 361-373.

63. Chirlian, L.E., *Atomic Charges Derived from Electrostatic Potentials - a Detailed Study*. Journal of Computational Chemistry, 1987. **8**(6): p. 894-905.

64. Born, M., *Volumes and hydration warmth of ions*, in *Z Phys*. 1920. p. 45-48.

65. Kirkwood, J.G., *Quantum statistics of almost classical assemblies*, in *Phys Rev*. 1934. p. 0116-0117.

66. Onsager, L., *Electric moments of molecules in liquids*, in *Journal of the American Chemical Society*. 1936. p. 1486-1493.

67. RINALDI, D., M.F. RUIZLOPEZ, and J.L. RIVAIL, *ABINITIO SCF CALCULATIONS ON ELECTROSTATICALLY SOLVATED MOLECULES USING A DEFORMABLE 3 AXES ELLIPSOIDAL CAVITY*, in *Journal of Chemical Physics*. 1983. p. 834-838.

68. RIVAIL, J.L., *COMPUTATION OF ELECTRON CORRELATION-EFFECTS IN A MOLECULE SOLVATED BY A CONTINUUM*, in *Cr Acad Sci Ii*. 1990. p. 307-311.

69. MIERTUS, S., E. SCROCCO, and J. TOMASI, *ELECTROSTATIC INTERACTION OF A SOLUTE WITH A CONTINUUM - A DIRECT UTILIZATION OF ABINITIO MOLECULAR POTENTIALS FOR THE PREVISION OF SOLVENT EFFECTS*, in *Chem Phys*. 1981. p. 117-129.

70. MIERTUS, S. and J. TOMASI, *APPROXIMATE EVALUATIONS OF THE ELECTROSTATIC FREE-ENERGY AND INTERNAL ENERGY CHANGES IN SOLUTION PROCESSES*, in *Chem Phys*. 1982. p. 239-245.

71. Baldridge, K.K. and A. Klamt, *First principles implementation of solvent effects without outlying charge error*, in *Journal of Chemical Physics*. 1997. p. 6622-6633.

72. Gregerson, L.N. and K.K. Baldridge, *Outlying charge, stability, efficiency, and algorithmic enhancements in the quantum-mechanical solvation method, COSab-GAMESS*. Helvetica Chimica Acta, 2003. **86**(12): p. 4112-4132.

73. BONE, S. and R. PETHIG, *DIELECTRIC STUDIES OF THE BINDING OF WATER TO LYSOZYME*, in *J Mol Biol*. 1982. p. 571-575.

74. BONE, S. and R. PETHIG, *DIELECTRIC STUDIES OF PROTEIN HYDRATION AND HYDRATION-INDUCED FLEXIBILITY*, in *J Mol Biol*. 1985. p. 323-326.

75. Schneider, G. and U. Fechner, *Computer-based de novo design of drug-like molecules*, in *Nat Rev Drug Discov*. 2005.

76. Taylor, R.D., P. Jewsbury, and J.W. Essex, *A review of protein-small molecule docking methods*, in *J Comput Aid Mol Des*. 2002. p. 151–166.

77. Schneider, G. and H. Bohm, *Virtual screening and fast automated docking methods*, in *Drug Discovery Today*. 2002. p. 64-70.

78. Fischer, H.E., *Einfluss der Configuration auf die Wirkung den*

Enzyme. Ber. Dtsch. Chem. Ges, 1894. **27**(2985).

79. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis*. Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(2): p. 98-104.

80. Uhrinova, S., *Structure of free MDM2 N-terminal domain reveals conformational adjustments that accompany p53-binding*. Journal of Molecular Biology, 2005. **350**(3): p. 587-598.

81. DESJARLAIS, R.L., et al., *DOCKING FLEXIBLE LIGANDS TO MACROMOLECULAR RECEPTORS BY MOLECULAR SHAPE*, in *J Med Chem*. 1986. p. 2149-2153.

82. LEACH, A.R. and I.D. KUNTZ, *CONFORMATIONAL-ANALYSIS OF FLEXIBLE LIGANDS IN MACROMOLECULAR RECEPTOR-SITES*, in *Journal of Computational Chemistry*. 1992. p. 730-748.

83. PAULING, L. and M. Delbrück, *THE NATURE OF THE INTERMOLECULAR FORCES OPERATIVE IN BIOLOGICAL PROCESSES*, in *Science*. 1940. p. 77-79.

84. KAUZMANN, W., *SOME FACTORS IN THE INTERPRETATION OF PROTEIN DENATURATION*, in *Adv Protein Chem*. 1959. p. 1-63.

85. Eisenhaber, F., *Hydrophobic regions on protein surfaces*, in *Perspect Drug Discov*. 1999. p. 27-42.

86. Chalikian, T.V., *Hydrophobic polar groups tendencies of as a major force in molecular recognition*, in *Biopolymers*. 2003. p. 492-496.

87. Franks, F., *Protein stability: the value of 'old literature'*, in *Biophys Chem*. 2002. p. 117-127.

88. Warshel, A., *Modeling electrostatic effects in proteins*. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2006. **1764**: p. 1647-1676.

89. Gouy, G.L., j. de phys, 1910. **9**(457).

90. Chapman, D.L., Philos. Mag, 1913. **25**(475).

91. RICHARDS, F.M., *INTERPRETATION OF PROTEIN STRUCTURES - TOTAL VOLUME, GROUP VOLUME DISTRIBUTIONS AND PACKING DENSITY*, in *J Mol Biol*. 1974. p. 1-14.

92. NICHOLLS, A. and B. HONIG, *A RAPID FINITE-DIFFERENCE ALGORITHM, UTILIZING SUCCESSIVE OVER-RELAXATION TO SOLVE THE POISSON-BOLTZMANN EQUATION*, in *Journal of Computational Chemistry*. 1991. p. 435-445.

93. HONIG, B. and A. NICHOLLS, *CLASSICAL ELECTROSTATICS IN BIOLOGY AND CHEMISTRY*, in *Science*. 1995. p. 1144-1149.

94. Briggs, J.M., et al., *University of Houston Brownian Dynamics Program User's Guide and Programmer's Manual Release 5.1*, in *University of Houston*. 1989.

95. Schutz, C.N. and A. Warshel, *What axe the dielectric "constants" of proteins and how to validate electrostatic models?*, in *Proteins*. 2001. p. 400-417.

96. Sharp, K.A., *Electrostatic Interactions in Macromolecules - Theory and Applications*. Annual Review of Biophysics and Biophysical Chemistry, 1990. **19**: p. 301-332.

97. Wagoner, J.A. and N.A. Baker, *Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and …*, in *PNAS*. 2006. p. 8331-8336.

98. Foster, I. and C. Kesselman, *Computational grids - Invited talk (Reprinted from The Grid: Blueprint for a new computing infrastructure, 1998)*, in *Vector and Parallel Processing - Vecpar 2000*. 2001. p. 3-37.

99. SETI@home, http://setiathome.ssl.berkeley.edu/.

100. DOCKING@home, http://docking.cis.udel.edu/.

101. Altintas, I., et al., *Kepler: An Extensible System for Design and Execution of Scientific Workflows*, in *ssdbm*. 2004. p. 423.

102. Abramson, D., J. Giddy, and L. Kotler, *High performance parametric modeling with Nimrod/G: killer application for the global grid?*, in *Parallel and Distributed Processing Symposium, 2000. IPDPS 2000. Proceedings. 14th International*. 2000. p. 520 - 528.

103. Williamson, J.R., *Induced fit in RNA-protein recognition*, in *Nat Struct Biol*. 2000. p. 834-837.

104. Walters, W.P., et al., *Babel - a Molecular Information Interchange Hub*. Abstracts of Papers of the American Chemical Society, 1994. **207**: p. 36-CINF.

105. Hooft, R.W.W., *Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures*. Proteins-Structure Function and Genetics, 1996. **26**(4): p. 363-376.

106. Nielsen, J.E., *Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations*. Proteins-Structure Function and Genetics, 2001. **43**(4): p. 403-412.

107. Nielsen, J.E., et al., *Improving macromolecular electrostatics calculations*. Protein Engineering, 1999. **12**(8): p. 657-662.

108. Signorini, G.F., et al., *Energetic fitness of histidine protonation states in PDB structures*, in *Journal of Physical Chemistry B*. 2004. p. 12252-12257.

109. Dolinsky TJ, N.J., McCammon JA, Baker NA., *PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations*. Nucleic Acids Research, 2004. **32 W665-W667**.

110. Sirius, http://sirius.sdsc.edu/.

111. Avbelj, F. and L. Fele, *Role of main-chain electrostatics, hydrophobic effect and side-chain conformational entropy in determining the secondary structure of proteins*, in *J Mol Biol*. 1998. p. 665-684.

112. Becke, A.D., *A New Mixing of Hartree-Fock and Local Density-Functional Theories*. Journal of Chemical Physics, 1993. **98**(2): p. 1372-1377.

113. Kim, K., *Comparison of Density-Functional and Mp2 Calculations on the Water Monomer and Dimer*. Journal of Physical Chemistry, 1994. **98**(40): p. 10089-10094.

114. Stephens, P.J., *Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields*. Journal of Physical Chemistry, 1994. **98**(45): p. 11623-11627.

115. Breneman, C.M. and K.B. Wiberg, *Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis*. Journal of Computational Chemistry, 1990. **11**(3): p. 361-373.

116. Gilson, M.K., *The Dielectric-Constant of a Folded Protein*, in *Biopolymers*. 1986. p. 2097-2119.

117. Wagoner, J. and N.A. Baker, *Solvation forces on biomolecular structures: A comparison of explicit solvent and Poisson-Boltzmann models*, in *Journal of Computational Chemistry*. 2004. p. 1623-1629.

118. Holst, M., N. Baker, and F. Wang, *Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples*, in *Journal of Computational Chemistry*. 2000. p. 1319-1342.

119. Baker, N., M. Holst, and F. Wang, *Adaptive multilevel finite element solution of the Poisson-Boltzmann equation II. Refinement at solvent-accessible surfaces in biomolecular systems*, in *Journal of Computational Chemistry*. 2000. p. 1343-1352.

120. Massova, I., *Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies*. Journal of the American Chemical Society, 1999. **121**(36): p. 8133-8143.

121. Kuhn, B. and P. Kollman, *A ligand that is predicted to bind better to avidin than biotin: insights from computational …*, in *J. Am. Chem. Soc*. 2000.

122. Baker, N.A., et al., *Electrostatics of nanosystems: Application to microtubules and the ribosome*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(18): p. 10037-10041.

123. Eisenberg, D., *Solvation Energy in Protein Folding and Binding*. Nature, 1986. **319**(6050): p. 199-203.

124. Lee, B., *Interpretation of Protein Structures - Estimation of Static Accessibility*. Journal of Molecular Biology, 1971. **55**(3): p. 379-&.

125. Chothia, C., *Hydrophobic Bonding and Accessible Surface-Area in Proteins*. Nature, 1974. **248**(5446): p. 338-339.

126. Wagoner, J.A., *Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(22): p. 8331-8336.

127. Pierotti, R.A., *Scaled Particle Theory of Aqueous and Non-Aqueous Solutions*. Chemical Reviews, 1976. **76**(6): p. 717-726.

128. Widom, B., *Critical-Point Thermodynamics of Fluids without Hole-Particle Symmetry*. Journal of Chemical Physics, 1973. **58**(2): p. 616-625.

129. Floris, F., *Evaluation of the Dispersion Contribution to the Solvation Energy - a Simple Computational Model in the Continuum Approximation*. Journal of Computational Chemistry, 1989. **10**(5): p. 616-627.

130. Weeks, J.D., *Perturbation Theory of Thermodynamic Properties of Simple Liquids*. Journal of Chemical Physics, 1971. **55**(11): p. 5422-+.

131. Weeks, J.D., *Role of Repulsive Forces in Determining Equilibrium Structure of Simple Liquids*. Journal of Chemical Physics, 1971. **54**(12): p. 5237-+.

132. Warshel, A., *Electrostatic Energy and Macromolecular Function*. Annual Review of Biophysics and Biophysical Chemistry, 1991. **20**: p. 267-298.

133. Nakamura, H., *A Theoretical-Study of the Dielectric-Constant of Protein*. Protein Engineering, 1988. **2**(3): p. 177-183.

134. Lund, M., B. Jonsson, and C.E. Woodward, *Implications of a high dielectric constant in proteins*, in *Journal of Chemical Physics*. 2007. p. 225103.

135. Emsley, J.o.h.n., *The Elements*. 1998. 304.

136. Abramson, D., et al., *A Flexible Grid Framework for Automatic Protein-Ligand Docking*, in *e-Science and Grid Computing*. 2006.

137. MASAMUNE, S., et al., *DOUBLE ASYMMETRIC-SYNTHESIS AND A NEW STRATEGY FOR STEREOCHEMICAL CONTROL IN ORGANIC-SYNTHESIS*, in *Angew Chem Int Edit*. 1985. p. 1-30.

138. Fong, S., *Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars*. Chemistry & Biology, 2000. **7**(11): p. 873-883.

139. Machajewski, T.D., *The catalytic asymmetric aldol reaction*. Angewandte Chemie-International Edition, 2000. **39**(8): p. 1352-1374.

140. Jiang, L., *De novo computational design of retro-aldol enzymes*. Science, 2008. **319**(5868): p. 1387-1391.

141. Zhong, G.F., *Catalytic enantioselective retro-aldol reactions: Kinetic resolution of beta-hydroxyketones with aldolase antibodies*. Angewandte Chemie-International Edition, 1998. **37**(18): p. 2481-2484.

142. TANOKURA, M., *H-1-NMR STUDY ON THE TAUTOMERISM OF THE IMIDAZOLE RING OF HISTIDINE-RESIDUES .2. MICROENVIRONMENTS OF HISTIDINE-12 AND HISTIDINE-119 OF BOVINE PANCREATIC RIBONUCLEASE-A*, in *Biochim Biophys Acta*. 1983. p. 586-596.

143. Fischer, P.M., *Peptide, peptidomimetic, and small-molecule antagonists of the p53-HDM2 protein-protein interaction*, in *Int J Pept Res Ther*. 2006. p. 3-19.

144. Murray, J.K. and S.H. Gellman, *Targeting protein-protein interactions: lessons from p53/MDM2*, in *Biopolymers*. 2007. p. 657-86.

145. Fasan, R., *Structure-activity studies in a family of beta-hairpin protein epitope mimetic inhibitors of the p53-HDM2 protein-protein interaction*. Chembiochem, 2006. **7**: p. 515-526.

146. Sakurai, K., *Crystallographic analysis of an 8-mer p53 peptide analogue complexed with MDM2*. Journal of the American Chemical Society, 2006. **128**(34): p. 11000-11001.

147. Massova, I. and P.A. Kollman, *Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies*, in *Journal of the American Chemical Society*. 1999. p. 8133-8143.

148. Kussie, P.H., et al., *Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain*, in *Science*. 1996. p. 948-53.

149. Fasan, R., et al., *Using a beta-hairpin to mimic an alpha-helix: cyclic peptidomimetic inhibitors of the p53-HDM2 protein-protein interaction*, in *Angew Chem Int Ed Engl*. 2004. p. 2109-12.

150. Garcia-Echeverria, C., et al., *… of potent antagonists of the interaction between human double minute 2 and tumor suppressor p53.*, in *J Med Chem*. 2000.

151. Portmann, S. and H.P. Lüthi, Chimia, 2000. **54**: p. 766.

152. Politzer, P., J.S. Murray, and M.C. Concha, *Halogen bonding and the design of new materials: organic bromides, chlorides and perhaps even fluorides as donors*, in *J Mol Model*. 2007. p. 643-650.

153. Seebeck, F.P., et al., *Stereoselectivity and expanded substrate scope of an engineered PLP-dependent aldolase*, in *Angew Chem Int Edit*. 2006. p. 6824-6826.

8 Appendix

8.1 Appendix 1: Estimation of the affinity between HDM2 receptor and 6 peptides along the rotation of both phenyl rings of phenylalanine 86 and 91 from the binding pocket.

Rotation of phenyl ring F91 and F86

F86	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91
	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
0	-18.00	-17.96	-17.99	-17.99	-18.01	-17.97	-17.95	-17.97	-18.01	-18.01	-18.08	-18.07	-18.03	-18.06	-18.09	-18.08	-18.05	-18.03	-18.05
10	-18.00	-17.97	-17.99	-17.98	-18.00	-17.98								-18.08	-18.10	-18.09	-18.08	-18.06	-18.07
20	-17.98	-17.94	-17.98	-17.97	-17.98	-17.96								-18.06	-18.09	-18.08	-18.08	-18.05	-18.06
30	-18.03	-17.99	-18.04	-18.03	-18.03	-18.02	-18.00						-18.15	-18.18	-18.20	-18.19	-18.17	-18.15	-18.15
40	-18.03	-17.99	-18.04	-18.03	-18.02	-18.01	-17.99	-18.01			-18.18	-18.17	-18.13	-18.17	-18.19	-18.18	-18.17	-18.15	-18.15
50	-18.03	-17.99	-18.05	-18.05	-18.05	-18.02	-18.00	-18.02	-18.05	-18.08	-18.18	-18.20	-18.15	-18.20	**-18.22**	-18.21	-18.19	-18.17	-18.17
60	-18.02	-17.98	-18.03	-18.04	-18.07	-18.02	-18.00	-18.02	-18.05	-18.07	-18.16	-18.18	-18.13	-18.18	-18.20	-18.19	-18.17	-18.15	-18.15
70	-18.02	-17.98	-18.04	-18.04	-18.07	-18.02	-18.01	-18.03	-18.06	-18.07	-18.18	-18.18	-18.14	-18.18	-18.20	-18.19	-18.17	-18.15	-18.15
80	-18.01	-17.97	-18.03	-18.03	-18.06	-18.01	-18.00	-18.02	-18.05	-18.06	-18.16	-18.17	-18.12	-18.17	-18.19	-18.18	-18.16	-18.13	-18.13
90	-18.02	-17.98	-18.04	-18.05	-18.06	-18.01	-18.00	-18.02	-18.05	-18.05	-18.18	-18.18	-18.14	-18.18	-18.21	-18.19	-18.17	-18.14	-18.14
100	-18.01	-17.98	-18.03	-18.05	-18.06	-18.01	-18.00	-18.02	-18.05	-18.05	-18.18	-18.19	-18.14	-18.19	-18.21	-18.19	-18.17	-18.14	-18.14
110	-18.01	-17.97	-18.03	-18.04	-18.07	-18.01	-18.00	-18.02	-18.05	-18.06	-18.18	-18.19	-18.14	-18.19	-18.21	-18.19	-18.17	-18.14	-18.14
120	-18.01	-17.97	-18.01	-18.03	-18.04	-17.99	-17.98	-18.01	-18.03	-18.04	-18.16	-18.17	-18.12	-18.17	-18.19	-18.17	-18.15	-18.12	-18.12
130	-18.03	-17.99	-18.03	-18.04	-18.05	-18.00	-17.99	-18.01	-18.04	-18.04	-18.16	-18.15	-18.13	-18.17	-18.19	-18.17	-18.17	-18.14	-18.14
140	-18.03	-17.99	-18.03	-18.04	-18.05	-17.99	-17.98	-18.00	-18.03	-18.04	-18.14	-18.15	-18.13	-18.17	-18.19	-18.18	-18.16	-18.13	-18.13
150	-18.00	-17.96	-18.00	-18.01	-18.01	-17.97	-17.96	-17.97	-18.01	-18.02					-18.17	-18.15	-18.13	-18.11	-18.11
160	-18.00	-17.96	-17.99	-17.99	-17.98	-17.95	-17.94	-17.95							-18.14	-18.12	-18.11	-18.08	-18.08
170	-18.02	-17.98	-18.00	-17.98	-18.00	-17.98	-17.96								-18.14	-18.13	-18.13	-18.11	-18.11
180	-18.00	-17.96	-17.99	-17.99	-18.00	-17.98	-17.96								-18.10	-18.09	-18.07	-18.02	-18.03

Figure: Binding energy between linear peptide with wild type tryptophan and HDM2 receptor from 1ycr crystal structure along rotation of both phenyl ring of F86 and F91.

Rotation of phenyl ring F91 and F86

F86	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91
	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
0	-18.84	-18.81	-18.84	-18.85	-18.88	-18.85	-18.84	-18.86	-18.88	-18.89	-18.95	-18.94	-18.89	-18.91	-18.94	-18.92	-18.89	-18.87	-18.88
10	-18.77	-18.73	-18.85	-18.86	-18.89	-18.88	-18.88	-18.91	-18.95	-18.94	**-19.05**	-19.04	-18.95	-18.99	-19.03	-19.00	-18.97	-18.92	-18.92
20	-18.84	-18.80	-18.86	-18.83	-18.86	-18.84								-18.94	-18.96	-18.94	-18.92	-18.89	-18.90
30	-18.80	-18.79	-18.83	-18.80	-18.82	-18.81								-18.91	-18.93	-18.91	-18.91	-18.87	-18.88
40	-18.82	-18.81	-18.88	-18.85	-18.86	-18.86	-18.89						-18.96	-18.99	-19.00	-18.98	-18.96	-18.93	-18.93
50	-18.81	-18.80	-18.86	-18.85	-18.85	-18.86	-18.88	-18.90			-19.03	-19.01	-18.95	-18.96	-18.97	-18.95	-18.94	-18.92	-18.92
60	-18.81	-18.80	-18.86	-18.85	-18.88	-18.87	-18.89	-18.91	-18.94	-18.95	-19.04	**-19.05**	-18.96	-18.99	-19.00	-18.97	-18.95	-18.92	-18.92
70	-18.79	-18.78	-18.85	-18.84	-18.89	-18.87	-18.90	-18.93	-18.95	-18.94	-19.03	-19.03	-18.95	-18.98	-18.98	-19.00	-18.97	-18.94	-18.91
80	-18.79	-18.78	-18.85	-18.85	-18.89	-18.88	-18.91	-18.93	-18.96	-18.95	**-19.05**	-19.04	-18.96	-18.98	-19.02	-18.99	-18.97	-18.94	-18.94
90	-18.78	-18.76	-18.83	-18.83	-18.88	-18.87	-18.90	-18.93	-18.95	-18.94	-19.04	-19.03	-18.94	-18.96	-19.00	-18.98	-18.97	-18.92	-18.92
100	-18.78	-18.76	-18.85	-18.86	-18.89	-18.87	-18.90	-18.92	-18.95	-18.93	**-19.05**	-19.04	-18.95	-18.97	-19.03	-19.00	-18.97	-18.93	-18.93
110	-18.75	**-18.71**	-18.80	-18.84	-18.89	-18.88	-18.89	-18.91	-18.93	-18.91	-19.02	-19.01	-18.94	-18.97	-18.99	-18.95	-18.91	-18.86	-18.86
120	-18.75	**-18.71**	-18.78	-18.82	-18.88	-18.87	-18.88	-18.90	-18.92	-18.90	-19.01	-19.00	-18.92	-18.95	-18.96	-18.93	-18.89	-18.85	-18.85
130	-18.77	-18.74	-18.80	-18.84	-18.89	-18.88	-18.89	-18.92	-18.93	-18.91	-19.01	-18.99	-18.93	-18.94	-18.96	-18.93	-18.91	-18.87	-18.87
140	-18.79	-18.76	-18.83	-18.86	-18.90	-18.89	-18.89	-18.92	-18.94	-18.91	-19.00	-18.99	-18.94	-18.96	-18.98	-18.95	-18.93	-18.89	-18.89
150	-18.77	-18.75	-18.82	-18.85	-18.87	-18.88	-18.87	-18.89	-18.91	-18.90					-18.97	-18.94	-18.92	-18.88	-18.88
160	-18.81	-18.78	-18.84	-18.85	-18.88	-18.86	-18.86	-18.87							-18.94	-18.93	-18.91	-18.86	-18.87
170	-18.85	-18.82	-18.86	-18.87	-18.90	-18.89	-18.88								-18.94	-18.93	-18.93	-18.90	-18.91
180	-18.85	-18.82	-18.86	-18.87	-18.89	-18.87	-18.86								-18.92	-18.91	-18.89	-18.87	-18.88

Figure: Binding energy between linear peptide with 6-chloro-tryptophan and HDM2 receptor from 1ycr crystal structure along rotation of both phenyl ring of F86 and F91.

Rotation of phenyl ring F91 and F86

F86	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91
	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
0	-18.78	-18.74	-18.76	-18.76	-18.78	-18.75	-18.73	-18.75	-18.79	-18.80	-18.89	-18.89	-18.83	-18.86	-18.90	-18.89	-18.86	-18.84	-18.84
10	-18.78	-18.73	-18.76	-18.75	-18.76	-18.74								-18.89	-18.91	-18.90	-18.89	-18.86	-18.87
20	-18.74	-18.69	-18.73	-18.71	-18.72	-18.71								-18.85	-18.88	-18.87	-18.88	-18.84	-18.84
30	-18.76	-18.71	-18.78	-18.77	-18.77	-18.76	-18.75						-18.88	-18.93	-18.96	-18.94	-18.92	-18.90	-18.90
40	-18.76	-18.71	-18.76	-18.76	-18.75	-18.75	-18.74	-18.76			-18.94	-18.93	-18.87	-18.91	-18.99	-18.97	-18.97	-18.89	-18.89
50	-18.75	-18.71	-18.76	-18.77	-18.78	-18.76	-18.75	-18.77	-18.81	-18.83	-18.95	-18.97	-18.88	-18.93	**-19.01**	-19.00	-18.98	-18.94	-18.89
60	-18.73	-18.69	-18.75	-18.75	-18.79	-18.76	-18.76	-18.78	-18.82	-18.82	-18.94	-18.96	-18.87	-18.92	-19.00	-18.98	-18.96	-18.93	-18.92
70	-18.73	-18.69	-18.75	-18.76	-18.80	-18.77	-18.77	-18.79	-18.83	-18.83	-18.96	-18.96	-18.87	-18.92	-18.99	-18.98	-18.96	-18.92	-18.92
80	-18.72	-18.67	-18.74	-18.75	-18.79	-18.77	-18.77	-18.79	-18.82	-18.82	-18.95	-18.95	-18.86	-18.90	-18.98	-18.97	-18.95	-18.92	-18.91
90	-18.73	-18.68	-18.75	-18.77	-18.79	-18.77	-18.77	-18.79	-18.82	-18.82	-18.97	-18.96	-18.87	-18.94	-19.00	-18.98	-18.96	-18.92	-18.91
100	-18.74	-18.67	-18.75	-18.78	-18.80	-18.78	-18.78	-18.80	-18.83	-18.83	-18.97	-18.97	-18.87	-18.95	-18.99	-18.98	-18.95	-18.91	-18.90
110	-18.71	-18.67	-18.74	-18.77	-18.81	-18.78	-18.78	-18.81	-18.84	-18.83	-18.97	-18.97	-18.88	-18.94	-18.97	-18.94	-18.92	-18.87	-18.86
120	-18.71	**-18.66**	-18.72	-18.74	-18.79	-18.77	-18.77	-18.80	-18.83	-18.82	-18.95	-18.95	-18.86	-18.91	-18.95	-18.92	-18.90	-18.86	-18.85
130	-18.73	-18.68	-18.73	-18.76	-18.80	-18.78	-18.78	-18.81	-18.83	-18.82	-18.95	-18.93	-18.87	-18.91	-18.94	-18.92	-18.91	-18.87	-18.87
140	-18.74	-18.69	-18.75	-18.77	-18.81	-18.78	-18.78	-18.80	-18.83	-18.82	-18.93	-18.93	-18.87	-18.90	-18.95	-18.93	-18.91	-18.88	-18.87
150	-18.71	-18.68	-18.73	-18.75	-18.77	-18.76	-18.76	-18.77	-18.80	-18.80					-18.93	-18.92	-18.90	-18.86	-18.86
160	-18.74	-18.70	-18.75	-18.75	-18.77	-18.75	-18.74	-18.75							-18.92	-18.90	-18.88	-18.85	-18.83
170	-18.78	-18.74	-18.77	-18.77	-18.80	-18.78	-18.76								-18.93	-18.92	-18.92	-18.86	-18.86
180	-18.79	-18.74	-18.77	-18.77	-18.78	-18.76	-18.75								-18.87	-18.86	-18.85	-18.82	-18.83

Figure: Binding energy between linear peptide with 6-methyl-tryptophan and HDM2 receptor from 1ycr crystal structure along rotation of both phenyl ring of F86 and F91.

Rotation of phenyl ring F91 and F86

	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
F86	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91
0	-17.75	-17.77	-17.79	-17.80	-17.80	-17.74	-17.74	-17.78	-17.79	-17.76	-17.79	-17.79	-17.81	-17.79	-17.73	-17.72	-17.75	-17.72	-17.76
10	-17.71	-17.72	-17.74	-17.75	-17.76	-17.70	-17.69	-17.73	-17.75	-17.71	-17.74	-17.74	-17.76	-17.74	-17.68	-17.67	-17.70	-17.67	-17.71
20	-17.73	-17.75	-17.77	-17.78	-17.78	-17.72	-17.71	-17.76	-17.77	-17.73	-17.76	-17.77	-17.79	-17.77	-17.70	-17.70	-17.72	-17.70	-17.73
30	-17.76	-17.77	-17.79	-17.80	-17.81	-17.74	-17.73	-17.78	-17.78	-17.75	-17.78	-17.79	-17.81	-17.80	-17.74	-17.73	-17.76	-17.73	-17.76
40	-17.73	-17.74	-17.77	-17.78	-17.78	-17.71	-17.71	-17.75	-17.75	-17.71	-17.75	-17.75	-17.78	-17.77	-17.71	-17.70	-17.73	-17.70	-17.73
50	-17.68	-17.70	-17.72	-17.73	-17.73	-17.67	-17.66	-17.70	-17.71	-17.67			-17.74	-17.72	-17.66	-17.66	-17.68	-17.66	-17.69
60	-17.73	-17.74	-17.77	-17.77	-17.78	-17.71	-17.71	-17.75							-17.71	-17.70	-17.73	-17.70	-17.73
70	-17.72	-17.73	-17.75	-17.76	-17.76	-17.70	-17.70									-17.69	-17.72	-17.69	-17.71
80	-17.75	-17.76	-17.78	-17.79	-17.79											-17.74	-17.76	-17.73	-17.75
90	-17.76	-17.76	-17.78	-17.79												-17.74	-17.76	-17.73	-17.75
100	-17.76	-17.77	-17.79													-17.75	-17.77	-17.74	-17.76
110	-17.76	-17.77	-17.79												-17.76	-17.75	-17.77	-17.74	-17.76
120	-17.76	-17.78	-17.80											-17.82	-17.76	-17.75	-17.77	-17.74	-17.76
130	-17.75	-17.76	-17.79											-17.80	-17.74	-17.73	-17.76	-17.72	-17.75
140	-17.76	-17.76	-17.78	-17.80									-17.82	-17.80	-17.74	-17.72	-17.75	-17.72	-17.75
150	-17.75	-17.76	-17.77	-17.79	-17.79							-17.79	-17.80	-17.78	-17.72	-17.71	-17.74	-17.71	-17.74
160	-17.75	-17.77	-17.79	-17.79	-17.80	-17.74					-17.79	-17.79	-17.81	-17.78	-17.72	-17.71	-17.76	-17.73	-17.76
170	-17.75	-17.76	-17.79	-17.80	-17.79	-17.73	-17.73	-17.77	-17.79	-17.75	-17.78	-17.78	-17.80	-17.78	-17.72	-17.71	-17.74	-17.72	-17.75
180	-17.74	-17.76	-17.78	-17.80	-17.80	-17.74	-17.73	-17.78	-17.79	-17.75	-17.78	-17.78	-17.80	-17.78	-17.72	-17.71	-17.74	-17.71	-17.75

Figure: Binding energy between cyclic peptide with wild type tryptophan and HDM2 receptor from 2axi crystal structure along rotation of both phenyl ring of F86 and F91.

Rotation of phenyl ring F91 and F86

	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
F86	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91
0	-18.70	-18.73	-18.76	-18.79	-18.81	-18.76	-18.78	-18.78	-18.77	-18.72	-18.74	-18.73	-18.74	-18.72	-18.65	-18.65	-18.68	-18.66	-18.71
10	-18.68	-18.71	-18.74	-18.77	-18.78	-18.73	-18.75	-18.75	-18.74	-18.69	-18.71	-18.70	-18.71	-18.69	-18.63	**-18.62**	-18.66	-18.64	-18.68
20	-18.73	-18.75	-18.79	-18.81	-18.83	-18.78	-18.79	-18.79	-18.78	-18.73	-18.75	-18.75	-18.76	-18.74	-18.68	-18.67	-18.71	-18.69	-18.73
30	-18.76	-18.79	-18.81	-18.84	-18.85	-18.80	-18.82	-18.81	-18.79	-18.75	-18.77	-18.76	-18.78	-18.77	-18.72	-18.71	-18.75	-18.73	-18.77
40	-18.80	-18.82	-18.85	-18.87	-18.88	-18.83	-18.84	-18.83	-18.81	-18.77	-18.79	-18.78	-18.80	-18.80	-18.75	-18.75	-18.78	-18.76	-18.80
50	-18.79	-18.81	-18.84	-18.86	-18.87	-18.82	-18.83	-18.82	-18.80	-18.76			-18.80	-18.80	-18.74	-18.74	-18.78	-18.76	-18.79
60	-18.88	-18.90	-18.93	-18.95	-18.96	-18.91	-18.92	-18.91							-18.84	-18.84	-18.87	-18.85	-18.89
70	-18.87	-18.89	-18.92	-18.94	-18.94	-18.89	-18.91									-18.83	-18.86	-18.84	-18.87
80	-18.90	-18.92	-18.94	-18.97	**-18.98**											-18.87	-18.90	-18.87	-18.91
90	-18.89	-18.90	-18.94	-18.96												-18.86	-18.89	-18.86	-18.89
100	-18.87	-18.89	-18.92													-18.85	-18.87	-18.85	-18.88
110	-18.84	-18.86	-18.89												-18.83	-18.82	-18.85	-18.82	-18.85
120	-18.81	-18.83	-18.86											-18.87	-18.80	-18.79	-18.82	-18.79	-18.82
130	-18.75	-18.77	-18.81											-18.81	-18.74	-18.72	-18.75	-18.73	-18.76
140	-18.75	-18.76	-18.79	-18.82									-18.82	-18.79	-18.72	-18.71	-18.73	-18.71	-18.74
150	-18.72	-18.74	-18.77	-18.80	-18.82							-18.77	-18.78	-18.75	-18.69	-18.67	-18.70	-18.68	-18.71
160	-18.68	-18.70	-18.74	-18.76	-18.78	-18.74					-18.73	-18.72	-18.73	-18.70	-18.64	-18.63	-18.68	-18.65	-18.69
170	-18.69	-18.71	-18.74	-18.77	-18.79	-18.75	-18.77	-18.77	-18.76	-18.71	-18.73	-18.72	-18.73	-18.70	-18.64	-18.63	-18.66	-18.64	-18.68
180	-18.70	-18.72	-18.76	-18.79	-18.80	-18.76	-18.78	-18.77	-18.76	-18.72	-18.73	-18.72	-18.73	-18.71	-18.65	-18.64	-18.67	-18.66	-18.70

Figure: Binding energy between cyclic peptide with 6-chloro-tryptophan and HDM2 receptor from 2axi crystal structure along rotation of both phenyl rings of F86 and F91.

Rotation of phenyl ring F91 and F86

	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180
F86	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91	F91
0	-18.64	-18.64	-18.65	-18.65	-18.65	-18.59	-18.60	-18.62	-18.63	-18.63	-18.67	-18.67	-18.70	-18.68	-18.62	-18.62	-18.65	-18.62	-18.65
10	-18.62	-18.62	-18.62	-18.62	-18.62	-18.56	-18.57	-18.59	-18.60	-18.60	-18.63	-18.64	-18.66	-18.64	-18.59	-18.59	-18.62	-18.59	-18.62
20	-18.64	-18.64	-18.65	-18.65	-18.65	-18.58	-18.60	-18.61	-18.62	-18.62	-18.66	-18.66	-18.68	-18.67	-18.62	-18.62	-18.65	-18.62	-18.65
30	-18.65	-18.65	-18.65	-18.65	-18.65	-18.58	-18.60	-18.61	-18.62	-18.62	-18.65	-18.66	-18.69	-18.68	-18.63	-18.63	-18.66	-18.63	-18.65
40	-18.67	-18.67	-18.68	-18.68	-18.67	-18.61	-18.62	-18.63	-18.64	-18.64	-18.67	-18.68	-18.71	-18.70	-18.65	-18.65	-18.68	-18.66	-18.68
50	-18.65	-18.65	-18.66	-18.65	-18.65	-18.58	-18.60	-18.61	-18.62	-18.62			-18.69	-18.68	-18.63	-18.63	-18.66	-18.63	-18.65
60	-18.73	-18.73	-18.74	-18.73	-18.73	-18.67	-18.68	-18.70							-18.72	-18.72	-18.75	-18.72	-18.74
70	-18.73	-18.72	-18.73	-18.72	-18.72	-18.66	-18.67									-18.72	-18.75	-18.71	-18.73
80	-18.77	-18.76	-18.77	-18.76	-18.76											-18.77	-18.80	-18.76	-18.78
90	-18.77	-18.77	-18.77	-18.77												-18.78	-18.80	-18.77	-18.78
100	-18.78	-18.77	-18.77													-18.78	-18.81	-18.77	-18.78
110	-18.77	-18.77	-18.77												-18.79	-18.78	-18.81	-18.77	-18.78
120	-18.75	-18.75	-18.75											**-18.83**	-18.77	-18.76	-18.79	-18.75	-18.76
130	-18.70	-18.70	-18.70											-18.77	-18.71	-18.70	-18.73	-18.69	-18.71
140	-18.69	-18.68	-18.68	-18.68									-18.77	-18.75	-18.69	-18.68	-18.71	-18.67	-18.69
150	-18.67	-18.67	-18.66	-18.66	-18.67							-18.72	-18.74	-18.72	-18.66	-18.66	-18.68	-18.65	-18.66
160	-18.65	-18.64	-18.65	-18.64	-18.64	-18.59					-18.69	-18.69	-18.71	-18.69	-18.64	-18.63	-18.67	-18.64	-18.66
170	-18.64	-18.64	-18.64	-18.64	-18.64	-18.58	-18.60	-18.62	-18.64	-18.64	-18.67	-18.67	-18.69	-18.67	-18.62	-18.62	-18.65	-18.62	-18.64
180	-18.64	-18.64	-18.65	-18.64	-18.64	-18.58	-18.60	-18.61	-18.63	-18.63	-18.66	-18.66	-18.68	-18.67	-18.62	-18.61	-18.65	-18.62	-18.64

Figure: Binding energy between cyclic peptide with 6-methyl-tryptophan and HDM2 receptor from 2axi crystal structure along rotation of both phenyl rings of F86 and F91.

8.2 Appendix 2: Binding energy of the complex HDM2 and the cyclic peptide with chlorine substituent on tryptophan 23, over translation of the peptide away from the receptor

2AXI-CLW	**Kcal/mol**	**Kcal/mol**	**Distance in Angstrom**		
PDIE	**ΔGelec**	**E TOT**	**(CLW)N-O(L32)**	**(CLW)Cl-C3(F86)**	**(CLW)Cl-C3(F91)**
2axi-1	10.8	-18.66	2.82	3.54	3.79
2axi-2	10.86	-18.24	2.91	3.76	3.86
2axi-3	11.81	-16.92	3.01	3.99	3.95
2axi-4	11.61	-16.38	3.2	4.32	4.11
2axi-5	11.05	-17.2	3.14	4.33	4.19
2axi-7	10.33	-18.16	3.21	4.2	3.78

Table: Total binding energy between the HDM2 receptor and Cyclic peptide substituted by the Chlorine atom along translation of the peptide.

2AXI-MEW	**Kcal/mol**	**Kcal/mol**	**Distance in Angstrom**		
PDIE	**ΔGelec**	**E TOT**	**(CLW)N-O(L32)**	**(CLW)Cl-C3(F86)**	**(CLW)Cl-C3(F91)**
2axi-0	10.84	-19.2	2.74	3.51	3.81
2axi-1	10.64	-19.27	2.7	3.63	3.91
2axi-2	10.62	-19.15	2.66	3.73	4.01
2axi-4	10.89	-18.71	2.63	3.85	4.11
2axi-5	10.85	-18.61	2.6	3.95	4.2
2axi-6	10.68	-18.49	2.57	4.11	4.34
2axi-7	10.65	-18.27	2.56	4.21	4.43
2axi-8	10.24	-18.37	2.54	5.51	4.69
2axi-9	11.21	-18.62	2.79	3.64	3.89
2axi-10	11.21	-18.5	2.74	3.77	4
2axi-11	10.64	-18.96	2.67	3.91	4.05
2axi-12	11.44	-18.16	2.79	3.86	3.99
2axi-13	7.34	-18.04	2.86	5.51	5.11
2axi-14	11.37	-18.18	2.88	3.73	3.82
2axi-15	11.57	-17.68	3	3.92	3.84
2axi-16	10.45	-17.42	3.52	4.62	4.02

Table: Total binding energy between the HDM2 receptor and Cyclic peptide substituted by the Methyl atom along translation of the peptide

2AXI-W	**Kcal/mol**	**Kcal/mol**	**Distance in Angstrom**		
PDIE	**ΔGelec**	**E TOT**	**(CLW)N-O(L32)**	**(CLW)Cl-C3(F86)**	**(CLW)Cl-C3(F91)**
2axi-0	11.1	-18.28	2.74	3.94	4.04
2axi-1	10.86	-18.36	2.73	4.06	4.14
2axi-2	10.96	-18.04	2.72	4.19	4.23
2axi-3	11.38	-17.38	2.71	4.33	4.35
2axi-4	11.42	-16.94	2.74	4.55	4.5
2axi-5	11.56	-16.15	2.79	4.83	4.74
2axi-6	10.93	-15.74	2.94	5.27	5.09
2axi-7	7.34	-11.55	4.03	7.15	6.75

Table: Total binding energy between the HDM2 receptor and Cyclic peptide with regular tryptophan residue along translation of the peptide

8.3 Appendix 3

8.3.1 Memory allocation in GAMESS

FM= Fast Memory

SETFM(MEMORY) => MEMORY = amount of memory in 8 byte words

VALFM(LOADFM) => LOADFM+1 = beginning address of memory that will be allocated

GOTFM(MEMMAX) => MEMMAX = amount of memory free

GETFM(NEED) => NEED = amount of memory needed in 8 byte words

RETFM => remove the memory

8.3.2 Implementation source code of a new dollar group integrated GAMESS and the reading of the data from APBS

```
C 3 NOV 06 - CA  - ADD THE NEW KEYWORD APBS IN GAMESS
C 13NOV 06 - CA  - ROUTINE APBS DX TO READ THE DX FILE FROM APBS


C*MODULE  APBS   *DECK APBSIN
      SUBROUTINE APBSIN
C
      IMPLICIT DOUBLE PRECISION (A-H, O-Z)
C
```

```
      LOGICAL IAPBS, UAPBS, TAPBS
C
      LOGICAL GOPARR,DSKWRK,MASWRK,ABEL
C
      INTEGER NPGRD
C
      COMMON /IOFILE/ IR,IW,IP,IS,IPK,IDAF,NAV,IODA(950)
      COMMON /PAR   / ME,MASTER,NPROC,IBTYP,IPTIM,GOPARR,DSKWRK,MASWRK
      COMMON /IAPBS / IAPBS, UAPBS, TAPBS
C
C     ----- set up the NAMELIST $APBS -----
C
      PARAMETER (NNAM=5)
      DIMENSION QNAM(NNAM),KQNAM(NNAM)
      DATA APBSWD/8HAPBS    /
      DATA QNAM/8HLAPBS  ,8HNAPROT  ,8HNACOMP  ,8HNALIG   ,8HNPGRD   /
      DATA KQNAM/1,1,1,1,1/


      UAPBS = .FALSE.
      IAPBS = .FALSE.
      TAPBS = .FALSE.
C
C DEFINE DEFAULT VALUES
C     LAPBS: SPECIFY THE TYPE OF COMPUTATION PROCEEDING WITH APBS
C          0 - SKIP THIS PROPERTY  (DEFAULT)
```

```
C          1 - BINDING ENERGY

C          2 - ...

C          3 - ...

     LAPBS=0

C

C    NAPROT: SPECIFY THE PROTEIN PQR FILE

C          0 - NO PROTEIN PQR FILE ASSOCIATED WITH THE COMPUTATION

C              (DEFAULT)

C          1 - A PROTEIN PQR FILE ASSOCIATED WITH THE COMPUTATION

     NAPROT=0

C

C    NACOMP: SPECIFY THE COMPLEX FILE

C          0 - DO NOT PUNCH A COMPLEX PQR FILE (DEFAULT)

C          1 - PUNCH A COMPLEX PQR FILE

     NACOMP=0

C

C    NALIG: SPECIFY THE LIGAND PQR

C          0 - DO NOT PUNCH A LIGAND PQR FILE

C          1 - PUNCH A LIGAND PQR FILE (DEFAULT)

     NALIG=1

C

C    NAINPT: SPECIFY THE APBS INPUT FILE

C         0 - DO NOT PUNCH AN APBS INPUT FILE (DEFAULT)

C          1 - PUNCH AN APBS INPUT FILE

C     NAINPT=0
```

```
C
C     NPGRD: SPECIFY THE NUMBER OF GRID POINT USE IN APBS
C          0 - 65*65*65=274625 (DEFAULT)
C          1 - PUNCH AN APBS INPUT FILE
      NPGRD=0
C
C READ THE  $APBS NAMELIST.
C
      JRET = 0
C
      CALL NAMEIO(IR,JRET,APBSWD,NNAM,QNAM,KQNAM,
     *           LAPBS, NAPROT, NACOMP, NALIG, NPGRD,
     *           0,0,0,0,0,
     *     0,0,0,0,0,   0,0,0,0,0,    0,0,0,0,0,
     *     0,0,0,0,0,   0,0,0,0,0,    0,0,0,0,0,   0,0,0,0,0
     *     0,0,0,0,0,   0,0,0,0,0,    0,0,0,0,0,   0,0,0,0,0)
C
      IF (JRET.EQ.1) RETURN
            IAPBS =.TRUE.
       CALL APBSDX(NPGRD)
C
      IF (JRET.EQ.2) THEN
       IF (MASWRK) WRITE(IW,9028) APBSWD
       CALL ABRT
      END IF
```

```
C
C  PRINT $APBS OPTIONS
C
      IF(MASWRK) WRITE(IW,9058) LAPBS,NAPROT,NACOMP,NALIG,NPGRD


C
      RETURN
C
 9028 FORMAT(1X,'ERROR IN $APBS INPUT - STOP ')
 9058 FORMAT(/5X,'$APBS OPTIONS'/5X,15(1H-)/
     *  5X,7HLAPBS =,I8,5X,8HNAPROT =,I8,5X,8HNACOMP =,I8/
     *  5X,7HNALIG =,I8,5X,7HNPGRD =,I8)


C
C     END OF APBSIN
C
      END


C*MODULE  APBS   *DECK APBSDX
      SUBROUTINE APBSDX(NPGRD)
C
C     READ DX FILE
C     DX NAME = INPUT NAME WITH DX EXTENSION
C
      IMPLICIT DOUBLE PRECISION(a-h,o-z)
```

```
C
      REAL xmin,ymin,zmin,hx,hy,hz,htemp
      REAL  CHG1,CHG2,CHG3
C
      CHARACTER*80 LINE,lword,lnx,lny,lnz
C
      INTEGER DX,i,j,k,ipnt,iline,kline,nx,ny,nz,icount
C
      PARAMETER (GRDPTS=65*65*65)
C
     COMMON /IOFILE/ IR,IW,IP,IS,IPK,IDAF,NAV,IODA(950)
     COMMON /FREFM2/ CARD,LCONT,LEOD,LEOC
C
     LOGICAL IAPBS, UAPBS, TAPBS
     COMMON /IAPBS / IAPBS, UAPBS, TAPBS
     COMMON /APBSFD/ GPAPBS,EAPBSC,EAPBSP,SEAPBS,COSAPBS(GRDPTS),
    *               CORAPBS(3,GRDPTS),NBCOOR


C
C Allocation memory declaration
C
      LOADFM = 0
      CALL GOTFM(MEMMAX)
      CALL VALFM(LOADFM)
      LCHG     = LOADFM + 1
```

```
      LCOOR     = LCHG   + 20

      LAST      = LCOOR  + 3*NPGRD

      NEED      = LAST - LOADFM - 1

      IF(NEED > MEMMAX) THEN

        write(7,*) 'max mem exceeded ',need,memmax

        CALL ABRT

      ENDIF

C

      CALL GETFM(NEED)

C

C Initialization variables

C

      icount=0

      IP = 7

      NBCOOR=0

 9020 FORMAT(A80)

 9030 FORMAT(4(E15.6),1X)

C

C     ----- OPEN THE DXFILE AND PUNCH FILES -----

C

      DX=39

      CALL SEQOPN(DX,'DXNAME', 'UNKNOWN',.TRUE., 'FORMATTED')

      write (7,*) '******* I open the file and npgrd=',NPGRD

C

C CHECK IF THE DX FILE IS EMPTY
```

```
C
      READ (DX,9020, IOSTAT=mycheck) LINE
      if (iost.GT.0) then
        CALL ABRT
      endif

 150   IF (MYCHECK.EQ.0) THEN
         write(7,*)'mycheck=',MYCHECK
         READ (DX,9020, IOSTAT=mycheck) LINE
         IF (index(LINE,'object').NE.0) THEN
         ELSE IF (index(LINE,'origin').NE.0) THEN
         write(7,*)'LINE: ', LINE
         ELSE IF (index(LINE,'delta').NE.0) THEN
         write(7,*)'LINE: ', LINE
         ELSE IF (index(LINE,'#').NE.0) THEN
         write(7,*)'LINE: ', LINE
         else
           NBCOOR=NBCOOR+1
         endif
         goto 150
       ENDIF

C       IF (MYCHECK.EQ.0) THEN
         NBCOOR=NBCOOR-1
         write(7,*)'nombre de lignes avec potential: ', NBCOOR
```

```
      write(7,*)'nombre de lignes du pqr: ',n
      REWIND (DX,iostat=ios)
      WRITE (7,FMT='(" $APBS   ")')
C
     do  iline=1,NBCOOR
      READ (DX,9020, IOSTAT=MYCHECK) LINE
C       WRITE (7,*) 'LINE= ', LINE
      if(index(LINE,'object 1').ne.0) then
        i=index(line,'counts')
        line=line(i+7:)
        i=index(line,' ')
        lnx=line(:i)
        read(lnx,'(I3)') nx
        write (7,*) 'NB UNIT GRID \n   X = ', nx
        line=line(i+1:)
        j=index(line,' ')
        lny=line(:j)
        read(lny,'(I3)') ny
        write (7,*) '  Y = ', ny
        line=line(j+1:)
        k=index(line,' ')
        read(line,'(I3)') nz
        write (7,*) '  Z = ', nz
        read(DX,'(A6,E14.6,E14.6,E14.6)')  word, xmin, ymin, zmin
        write (7,*) 'ORGIN'
```

```
            write (7,*) '  xmin = ', xmin
            write (7,*) '  ymin = ', ymin
            write (7,*) '  zmin = ', zmin
            read(DX,'(A5,3(E13.6))') word, hx
            write (7,*) 'SPACE GRID \n   x = ', hx
            read(DX,'(A5,3(E13.6))') word, htemp, hy
            write (7,*) '  y = ', hy
            read(DX,'(A5,3(E13.6))') lword, htemp, htemp, hz
            write (7,*) '  z = ', hz
          endif
C
         if   (index(LINE,'object 3').ne.0) then
          i = index(LINE,'data follows')
          write (7,*) 'NB GRID POINT  = ', line(47:i-1)
          LINE=LINE(47:i-1)
          READ(LINE,'(I15)') ipnt
          write(7,*)'ipnt = ', ipnt
         endif
       enddo
C
C  READ POTENTIAL
C
       do kline=1,NBCOOR
C         write(7,*)'icount=', icount
        if  (icount.LT.NPGRD) then
```

```
          READ (DX,'(3(E13.6,1X)))')   CHG1, CHG2, CHG3
          icount=icount+1
          COSAPBS(icount)=(CHG1/27.212)*100
          icount=icount+1
          COSAPBS(icount)=(CHG2/27.212)*100
          icount=icount+1
          COSAPBS(icount)=(CHG3/27.212)*100
          write(7,*)'***pot1,pot2,pot3:', CHG1, CHG2, CHG3
          write(7,*)'***read potential at line=', icount
        endif
C         write(7,*)'sortie de condition nb:', icount
       enddo
       write(7,*)'***read potential done - exit loop'
C
C
      CALL SEQCLO(DX,'DXNAME', 'OLD',.TRUE., 'FORMATTED')
      write(7,*)'*** close dx file'
C
C READ COORDINATES
C
       icount=1
       do 50 i=1,2
         CORAPBS(1,i) = xmin + hx*(i-1)
         do 50 j=1,2
           CORAPBS(2,j) = ymin + hy*(j-1)
```

```
            do 50 k=1,2
              CORAPBS(3,k) = zmin + hz*(k-1)
C              write(7,*)'corapbs(1,',i,')=',CORAPBS(1,i)
C               write(7,*)'corapbs(2,',j,')=',CORAPBS(2,j)
C               write(7,*)'corapbs(3',k,')=',CORAPBS(3,k)
C               icount=icount+1
 50     continue
C
       DO  IJK=1,2
       WRITE (7,9030) CORAPBS(1,IJK),CORAPBS(2,IJK),CORAPBS(3,IJK),
     *   COSAPBS(IJK)
       ENDDO
C
C      ENDIF
C
       CALL RETFM(NEED)
C
       WRITE(7,FMT='(" $END   ")' )
C
 9028 FORMAT(1X,'**** ERROR DX FILE MISSING OR EMPTY - STOP ****')
C
C      END OF APBSDX
C
      END
```